Durham E-Theses

Cognitive and neural mechanisms underlying auditory verbal hallucinations in a non-clinical sample

MOSELEY, PETER,WILLIAM

How to cite:

Use policy

Cognitive and neural mechanisms underlying auditory verbal hallucinations in

a non-clinical sample

Peter William Moseley, B.Sc., M.Sc.

A thesis submitted for the degree of Doctor of Philosophy in the Department of

Psychology at Durham University

December 2014

Table of contents

List of figures

Declaration

I confirm that no part of the material presented in this thesis has previously been submitted for a degree in this or in any other institution. If material has been generated through joint work, this has been indicated where appropriate. All other sources have been referenced, and quotations suitably indicated.

Statement of copyright

The copyright of this thesis rests with the author. No quotation from it should be published without the author's prior written consent and information derived from it should be acknowledged.

Acknowledgements

A massive thank you to my supervisors, Prof. Charles Fernyhough and Dr. Amanda Ellison. As well as securing funding for the PhD, they have both been incredibly supportive throughout the last three years, whilst also encouraging me to take my PhD in my own directions, for which I am extremely grateful.

I would also like to thank all the members of the Hearing the Voice research project, for all the inspiring conversations I have been a part of at Voice Club. Special thanks go to Dr. Ben Alderson-Day (for many excellent conversations, as well as many heated games of table tennis) and Dr. Sam Wilkinson (who lent his voice, both metaphorically to provide 'conceptual hygiene', and physically to the stimuli in Chapter 6). Dr. David Smailes, as well as helping with experimental design in Chapters 3 and 4, made my PhD much more enjoyable through our many, many conversations about psychology, neuroscience and football.

Thanks also to Dr. Judith Ford and Dr. Daniel Mathalon, who were extremely welcoming when I visited their lab in San Francisco, for one month during the PhD. I am also grateful to Dr. Jon Simons and the Memory Lab at Cambridge, who gave me some great feedback on plans for the study reported in Chapter 6.

My office-mates in Room 70 in the Durham Psychology department also deserve thanks for putting up with my rants, panics and ramblings. In particular, thanks to

9

Yan Birch, Kayleigh Carr, Alice Chadwick, Cath Grainger, Emma Grisdale, Hannah Harvey, Sophie Hodgetts, Katharina Kaduk, Emma Lough, and Anna Peel, who have the most experience of my shockingly high coffee intake. Thanks also to Norma Twomey, Elaine Stanton and Richard Stock, for always being willing to help with any technical or administrative problems.

I owe a big thank you to all my friends and family who have been there for me. In particular, Tony and Lesley Foster have provided an enormous amount of help over the years, for which I am extremely thankful. And I wouldn't have made it this far without my Mum and Dad, who have always supported me, whatever direction I wanted to take.

Last, but by no means least, I'd like to thank Stephanie Foster. I wouldn't have made it through a BSc, let alone an MSc and PhD, without her love and constant support.

Thesis abstract

Auditory verbal hallucinations (AVHs) are the experience of hearing a voice in the absence of any speaker. Cognitive models of AVHs have suggested that they may occur when an internal mental event, such as inner speech, is misattributed to an external source. This has variously been explained by reference to biases in self-monitoring, source monitoring, or reality discrimination processes. Evidence suggests that, mechanistically, this may be related to atypical functioning of a forward model system which usually predicts the outcome of self-generated actions, attenuating activity in sensory cortices to the resulting perceptual input. At a higher level, excessive vividness and low cognitive effort associated with internal mental events may be associated with external misattributions of inner speech. Chapter 1 reviews inner speech models of AVH, as well as recent attempts to reduce the frequency of AVHs using neurostimulation. Chapter 2 then provides a methodological overview of techniques used in this thesis.

The first two empirical studies presented in this thesis, in Chapters 3 and 4, explore the cognitive mechanisms underlying AVHs by investigating the associations between self-reported hallucination-proneness and phenomenology of inner speech, and performance on source monitoring and self-monitoring tasks, in a non-clinical, student sample. The results indicated that hallucination-prone participants were more likely to misattribute self-generated auditory verbal imagery, both when instructed to generate imagery and when they retrospectively reported using imagery. Regression analysis also indicated that a tendency to use dialogic inner speech, biased performance on reality discrimination and self-monitoring tasks, and a tendency to perceive meaning in jumbled speech independently predicted hallucination-proneness.

The studies presented in Chapters 5 and 6 investigated the neural basis of performance on auditory signal detection and source monitoring tasks using transcranial direct current stimulation (tDCS). Results indicated that modulating activity in the superior temporal gyrus/temporoparietal junction (STG/TPJ) affected the number of false perceptions on the signal detection task. However, stimulation to the left STG or medial prefrontal cortex did not affect performance on a source monitoring task. These results indicate that different cortical regions may be involved in the two tasks, and hence that they may reflect different aspects of how self-generated actions are experienced as such. Together, the four experimental chapters 1) provide evidence for inner speech accounts of AVH, 2) indicate the need for a more complex account of self-monitoring and reality discrimination in which both are seen as independent predictors of AVHs, and 3) suggest that the left STG plays a key role in reality discrimination, but less so in source monitoring tasks (at least in the encoding stage). The thesis concludes with a general discussion of these issues, and recommendations for future research.

Note on publications included in this thesis

At the time this thesis was submitted for examination, two of the chapters had been published.

Chapter 1:

Moseley, P., Fernyhough, C., & Ellison, A. (2013). Auditory verbal hallucinations as atypical inner speech monitoring, and the potential of neurostimulation as a treatment option. *Neuroscience and Biobehavioral Reviews, 37*(10), 2794-2805.

Chapter 5:

Moseley, P., Fernyhough, C., & Ellison, A. (2014). The role of the superior temporal lobe in auditory false perceptions: a transcranial direct current stimulation study. *Neuropsychologia, 62*, 202-208.

These two chapters are presented as they were submitted, although the referencing has been altered to be consistent throughout the thesis, and American English spelling has been altered to British English.

The experimental design used for the empirical studies presented in Chapters 3 and 4 was completed in conjunction with Dr. David Smailes. All data collection, analysis and writing were carried out by me.

Chapter 1

Auditory verbal hallucinations as atypical inner speech monitoring, and the potential of neurostimulation as a treatment option

Abstract

Auditory verbal hallucinations (AVHs) are the experience of hearing voices in the absence of any speaker, often associated with a schizophrenia diagnosis. Prominent cognitive models of AVHs suggest they may be the result of inner speech being misattributed to an external or non-self source, due to atypical self- or reality monitoring. These arguments are supported by studies showing that people experiencing AVHs often show an externalising bias during monitoring tasks, and neuroimaging evidence which implicates superior temporal brain regions, both during AVHs and during tasks that measure verbal self-monitoring performance. Recently, efficacy of noninvasive neurostimulation techniques as a treatment option for AVHs has been tested. Meta-analyses show a moderate effect size in reduction of AVH frequency, but there has been little attempt to explain the therapeutic effect of neurostimulation in relation to existing cognitive models. This article reviews inner speech models of AVHs, and argues that a possible explanation for reduction in frequency following treatment may be modulation of activity in the brain regions involving the monitoring of inner speech.

1. Introduction

Auditory verbal hallucinations (AVHs) are the phenomenon of hearing voices in the absence of any speaker, and are experienced by around 60-80% of people diagnosed with schizophrenia (Sartorius et al., 1986). Some studies also report that they are experienced by between 1.5% and 3% of the general population (Tien, 1991), highlighting that the experience is not always pathological, though estimates greatly vary between sources on this matter (Beavan, Read, & Cartwright, 2011). Despite the prevalence of this experience, surprisingly little is known about the cognitive and neural mechanisms underlying AVHs, and they may be refractory to current treatment options in around 25% of cases (Shergill, Murray, & McGuire, 1998).

A recent review by Sommer et al. (2012) suggested that antipsychotic medication such as olanzapine, amisulpride, ziprasidone or quetiapine may be the most efficacious treatment option for AVHs in schizophrenia, while clozapine should only be used in the event that these are unsuccessful. Anti-psychotic medication tends to block D_2-receptors in the brain, leading to hypotheses emphasizing the importance of dopamine pathways in the creation of psychotic experiences (Carlsson, 1978; Farde, 1997). However, it is well known that antipsychotic medication often causes undesirable side effects, such as weight gain and sedation (Buchanan et al., 2010). Therefore, cognitive behavioural therapy (CBT) is often used, either as an adjunctive or as an alternative treatment. The aim of CBT is to change the appraisal of the hallucination, in a collaborative effort between the patient and therapist; the patient is encouraged to take an active part in the therapy, for example, by examining evidence for and against distressing beliefs, and testing explanations for unusual experiences in real world situations (Jones, Hacker, Cormac, Meaden, & Irvine, 2012). One meta-analysis reported an effect size of 0.4 for a reduction in positive

symptoms of schizophrenia (Wykes, Steel, Everitt, & Tarrier, 2008), although this does not tell us specifically about CBT's efficacy in treating AVHs. These studies are also confounded by whether the patients included were taking anti-psychotic medication at the time of therapy; it is difficult to know whether any effects were due to the use of CBT alone.

The search for new treatment options for AVHs has led to the testing of the efficacy of noninvasive neurostimulation techniques in the treatment of AVHs. Although results have not been conclusive, repetitive pulse transcranial magnetic stimulation (rTMS) was recently labelled as "potentially useful" in a summary of available treatment options (Sommer et al., 2012, p. 7), and recent research has used transcranial direct current stimulation (tDCS), with promising results (Brunelin et al., 2012). Additionally, neurostimulation techniques, if indeed efficacious, have the potential to tell us much about the cognitive and neural mechanisms underlying AVHs, by targeting specific brain regions thought to be involved in the experience (although it also affects brain regions other than those directly underneath the stimulating coil or electrode; e.g., Kindler et al., 2013). There has so far been little attempt to explain the therapeutic effects of neurostimulation (if not a placebo effect) in relation to pre-existing cognitive or neuroscientific models of AVHs.

The most popular cognitive theory of AVHs is arguably that many are the result of internal cognitive events, such as inner speech, being misattributed to an external or alien source (Waters et al., 2012). Various models have suggested that this could be due to a specific deficit in the monitoring of one's own actions, known as self-monitoring (Frith, 1992), and/or due to a bias towards labelling internal mental events as externally produced under conditions of ambiguity, known as a bias in reality monitoring (Bentall & Slade, 1985). Evidence from neuroimaging suggests

that monitoring of one's own speech, overt or covert, is related to activity in auditory cortical regions such as the lateral temporal lobe, including the superior temporal gyri (STG), a brain area that includes both primary and secondary auditory cortices (Allen et al., 2007; McGuire, Silbersweig, & Frith, 1996). This corresponds well to 'symptom-capture' studies of AVHs, in which similar areas are often implicated (Allen, Larøi, McGuire, & Aleman, 2008). rTMS treatment is usually targeted at the left temporoparietal junction (TPJ), an area adjacent to, and with high levels of connectivity to, primary and secondary auditory cortex (Kindler et al., 2012). Therefore, it is possible that neurostimulation treatment affects brain regions involved in verbal self- or reality monitoring.

This review will discuss models that implicate atypical monitoring of inner speech, as well as the evidence surrounding the efficacy of neurostimulation as a treatment for AVHs, and the possible cognitive and neural mechanisms behind the therapeutic effect.

2. Auditory verbal hallucinations as the result of misattributed inner speech

Prominent models of AVHs have suggested that the experiences arise when an internal mental event is misattributed to an external or non-self source. For example, Frith (1992) suggests that, if inner speech is not recognized as self-initiated, it may be experienced as an AVH. Many models have assumed that the raw material of AVHs is a kind of inner speech (Bentall, 2003; Fernyhough, 2004), although definitions of inner speech have varied, from simply "thinking in words" (McGuire et al., 1995, p. 596) to "the overlapping region of thought and speech" (Jones &

Fernyhough, 2007a, p. 148), the latter of which highlights that not all thought processes necessarily take place as inner speech.

Perhaps the most compelling evidence that the raw material of AVHs is misattributed inner speech comes from studies that have used electromyography (EMG) to show subvocalization (tiny movements of the vocal musculature which occur during inner speech; Gould, 1948; Inouye & Shimizu, 1970; McGuigan, 1966) whilst patients experience AVHs. In one case, the subvocalizations were amplified into intelligible speech which corresponded well to the contents of the AVH (Green & Preston, 1981), and some AVHs have been shown to be less frequent when patients explicitly vocalized competing utterances, for example humming (Green & Kinsbourne, 1990). Further evidence from neuroimaging studies suggests that similar cortical areas are active during inner speech as during AVHs. For example, during auditory verbal imagery, Shergill et al. (2001) found activation in the left superior temporal gyrus (including Wernicke's area) and the left inferior frontal gyrus (Broca's area), as well as in the supplementary motor area (SMA) and insula. These findings concord fairly well with other inner speech functional neuroimaging studies (Friedman et al., 1998; McGuire et al., 1996). Raij & Riekki (2012) showed that the main difference between neural activation during AVHs and during imagining speech was that AVHs showed less activation in the SMA, otherwise implying that similar areas were recruited for imagining speech and AVHs. The functional localisation of inner speech has also been studied using single pulse TMS: Aziz-Zadeh, Cattaneo, Rochat and Rizzolatti (2005) were able to induce 'covert speech arrest' by stimulating either motor or non-motor language areas in the inferior frontal gyrus (IFG) in the left hemisphere, but not right hemispheric non-motor language areas.

In contrast, however, some have argued that left hemisphere language sites are not integral to the experience of AVHs. An fMRI study using a sample of 24 hallucinating patients concluded that the right homologue of Broca's area (IFG) and the right superior temporal gyrus, as well as the bilateral insula and anterior cingulate gyri, were most active during AVHs (Sommer et al., 2008). Vercammen, Knegtering, den Boer, Liemburg, & Aleman (2010) have also shown that functional connectivity of the left temporoparietal junction (TPJ) with the right homologue of Broca's area is reduced in patients who reported AVHs. These findings may be interpreted as discordant with the inner speech theory of AVHs, especially in light of Aziz-Zadeh et al.'s findings, which indicate that non-motor language areas in the right hemisphere are *not* causally involved in the production of inner speech. However, there are a number of possible explanations for right hemisphere involvement in AVHs. Vercammen et al. argue that inner speech generated by the right hemisphere may consist of short sentences, with negative or derogatory content, which seems to fit with phenomenological accounts of AVHs. It may simply be that the type of inner speech elicited by Aziz-Zadeh et al. did not recruit right hemisphere language areas. Alternatively, right-sided language areas could be involved in the contextualisation of AVHs (influencing emotional valence and attentional salience, for example). This suggestion would fit with findings that implicate right hemispheric activation in emotional prosody comprehension (Alba-Ferrara, Ellison, & Mitchell, 2012; Alba-Ferrara, Fernyhough, Weis, Mitchell, & Hausmann, 2012). Superior temporal regions of the right hemisphere are also important in processing aspects of speech such as pitch (Lattner, Meyer, & Friederici, 2005).

Alternatively, the right temporoparietal junction (rTPJ) has been implicated in theory of mind tasks (Young, Camprodon, Hauser, Pascual-Leone, & Saxe, 2010; Young, Dodell-Feder, & Saxe, 2010) and it has also been argued that the inferior parietal lobule (immediately adjacent to Wernicke's area and its right homologue), particularly on the right side, is important for feelings of self-agency (Jardri et al., 2007), leaving open the possibility that right-sided activation in AVHs is a result of the utilisation of some form of perspective taking mechanism (a possibility returned to later in this section). Inconsistent neuroimaging findings in relation to the lateralisation of AVHs may reflect the varying phenomenology of the experience, and it is likely that not all AVHs can be linked to inner speech (Jones, 2010).

Hoffman, Fernandez, Pittman & Hampson (2011) have argued that a better model to explain AVHs involves the surfacing of 'unbidden thoughts' into consciousness, through a hyperconnected corticostriatal loop involving Wernicke's area and its right homologue, the left inferior frontal gyrus, and the putamen bilaterally. This model also specifies that a possible reason for the experience of AVHs as another person's voice is linked to the activation of right-sided temporal areas. Although different in its details, this model is not incompatible with the typical view of inner speech as the raw material of AVHs, additionally emphasizing the importance of subcortical structures such as the putamen in conscious experience. The putamen is crucial in the initiation of language representations (Price, 2010), and Hoffman et al. argue that hyperconnectivity of the putamen with temporal and frontal areas represents an overabundance of language representations reaching temporal cortices. It is not immediately clear, though, why these language representations might be experienced as hallucinatory and as external to the self. If anything, the differences between the

unbidden thoughts model and inner speech models emphasize the need for a better understanding of the phenomenology of what we are referring to as 'inner speech'.

It may be that the differential findings of inner speech and AVH neuroimaging studies are in fact due to the type of task used to elicit inner speech. For example, many of the aforementioned studies have simply asked participants to repeat cued sentences in their heads, whilst Aziz-Zadeh et al. (2005) inferred covert speech arrest by observing an increase in reaction time in a syllable counting task. Although these tasks undoubtedly elicit some form of inner speech, their validity in relation to the kinds of inner speech that we experience in real life, or that may be related to the experience of AVHs, may be tenuous. For example, these forms of elicited inner speech may lack spontaneity and the phenomenal experience of an inner dialogue (see below). Future inner speech neuroimaging studies would therefore do well to utilize tasks that may elicit more realistic inner speech, as discussed in Section 6.

Inner speech theories of AVHs, though, have been criticized for not explaining the phenomenological aspects of AVHs. For example, AVHs are usually experienced as non-self generated and usually (but not always) located in external space. Furthermore, most hallucinations take the form of another person's voice, often giving commands or commenting on actions of the person, and usually being experienced as 'alien' to the self (Nayani & David, 1996). This does not seem to correspond to what most would associate with 'thinking in words', and the negative and often derogatory content of AVHs would also seem to contrast with this idea. One study reported no phenomenological difference between the inner speech of hallucinating patients diagnosed with schizophrenia and a control group of participants who did not hear voices (Langdon, Jones, Connaughton, & Fernyhough, 2009), whereas one might expect to find differences in, for example, the tendency to

21

represent others in inner speech (although, interestingly, the two questionnaire items which approached a significant effect were related to experiencing inner speech as a dialogic exchange). Also, an early neuroimaging study found no difference in brain activation between hallucinating patients and healthy controls during inner speech (McGuire, et al., 1995).

Fernyhough (1996, 2004) has argued that inner speech is fundamentally dialogic in nature, or 'shot through' with other voices. This is a logical extension of Vygotsky's (1934/1987) argument that inner speech is the result of the internalisation of external dialogues during psychological development. If true, it follows from Vygotsky's ideas that typical inner speech may consist of a dialogue, often including voices other than the person's own. One aspect of inner speech that has been shown to differ between hallucination-prone and non-hallucination prone healthy individuals is in fact self-reported propensity to use dialogic inner speech (McCarthy-Jones & Fernyhough, 2011), and the inner speech neuroimaging study by McGuire et al. did find differential activation between hallucinating patients and controls when participants were asked to imagine someone else's voice. This would seem to explain why most AVHs are experienced as a voice other than the person's own (McCarthy-Jones et al., 2012; Nayani & David, 1996): the atypical component of AVHs is not that they are experienced as someone else's voice, but instead that they are experienced as alien and/or external to the self (Jones & Fernyhough, 2007a). As already mentioned, the observation of right hemispheric activity during AVHs may in fact reflect engagement of a type of perspective-taking mechanism, which would be integral to the dialogicality of inner speech. This is backed up by the involvement of right temporal lobe involvement in theory of mind tasks (Young, Camprodon, Hauser, Pascual-Leone, & Saxe, 2010; Young, Dodell-Feder, & Saxe, 2010).

It therefore seems that the inner speech model is a good fit for at least some types of AVHs, although further research to elicit a better proxy for inner speech, and research that studies the neural correlates of different phenomenological types of AVHs (and inner speech), is needed before firm conclusions can be drawn.

3. Why do people who hear voices experience inner speech as alien?

If AVHs are indeed the result of inner speech being misattributed to an external/non-self source, it may follow that a mechanism that usually distinguishes between internally and externally produced stimuli is disrupted. This concept has variously been termed self-monitoring, source monitoring, or reality monitoring (Bentall, 1990; Frith, 1992). These terms have often been used interchangeably in the literature, or simply grouped under the umbrella term 'source monitoring'. In general, self-monitoring has tended to refer to the ability to monitor the planning and executing of actions (with inner speech being seen as a motor act), and has been associated with tasks requiring participants to monitor self-made actions or vocalisations (Frith, 1992). Meanwhile, source and reality monitoring have tended to be defined as the ability to distinguish between internal and external events, and have been associated with tasks requiring participants to recall whether a remembered item was produced by themselves or the experimenter (source memory) or signal detection tasks requiring participants to decide whether a voice is present in white noise or not (Bentall & Slade, 1985; Johnson, Hashtroudi, & Lindsay, 1993). Here, the terms will be used as described above, but the term 'monitoring' will also be used as an umbrella term to cover all of these concepts. (See Table 1.1 for a summary of cognitive tasks discussed in this review.)

23

Early self-monitoring studies measured schizophrenia patients' ability to monitor their own actions by using a simple joystick task in which participants had to monitor errors without feedback (Frith & Done, 1989). It was shown that those diagnosed with schizophrenia were worse at this form of monitoring than healthy controls. Neuroimaging studies investigating self-monitoring of speech specifically implicated a network of brain areas, involving the lateral temporal cortex bilaterally (consistent with theories that implicate similar brain regions in monitoring both internally and externally produced speech), as well as left inferior frontal cortices and hippocampal formations (McGuire, Silbersweig, & Frith, 1996). Following from early self-monitoring studies, one prominent theory has suggested that an internal forward model is disrupted in those that experience AVHs. This theory explains the feeling of agency that accompanies motor actions by postulating a system that uses the predicted consequences of actions to label events as self- or other-generated. Importantly, this theory relies on inner speech being seen as a covert motor action (Jones & Fernyhough, 2007b), which is supported by the aforementioned subvocalization research. A forward model account of self-monitoring and AVHs argues that when a motor plan is first created, an 'efference copy' or 'corollary discharge' of the plan is sent to sensory areas to 'warn' them that the action is about to occur (Ford & Mathalon, 2005). If the planned action then occurs (and appropriate sensory information is received as reafference), the event is labelled as self-generated (Seal, Aleman, & McGuire, 2004; Wolpert, Ghahramani, & Jordan, 1995). Some models specify that the efference copy will dampen activity in the appropriate sensory area, to label the percept as self-generated (Ford & Mathalon, 2005; Whitford, Ford, Mathalon, Kubicki, & Shenton, 2010). Applied to AVHs, this would

mean that an efference copy of the inner speech motor act has not reached auditory cortical areas (i.e., Wernicke's area, and the STG more generally).

The forward model account of self-monitoring has received support from tasks that require participants to discriminate between distorted voices (lowered either 3 or 6 semitones) that could be their own or someone else's. Here, a voice is immediately fed back to them through headphones when participants speak, and they are required to respond whether they think the voice is their own or not (McGuire, Silbersweig, & Frith, 1996). Studies utilising this task have shown that patients with AVHs are worse at making the self/other judgement correctly (Johns et al., 2001). Evidence of a global self-monitoring deficit in schizophrenia also comes from studies which show that patients with AVHs do not show a difference between the tickle sensation evoked by others and by themselves, when both healthy controls and patients without AVHs do (Blakemore, Smith, Steel, Johnstone, & Frith, 2000). Interpreted in light of the forward model theory of AVHs, typical individuals may not be able to tickle themselves because the corresponding sensory cortical areas are dampened when the efference copy of the motor plan successfully reaches it, whereas this may not be the case in hallucinating patients. These findings can therefore be seen to support the idea of a disrupted forward model self-monitoring system in those experiencing AVHs.

The neural instantiation of the efference copy has been postulated to be a dampened N1 event-related potential (ERP) during self-produced speech in comparison to other-produced speech in healthy controls, but not in patients with schizophrenia (Ford et al., 2001). Magnetoencephalography (MEG), the magnetic counterpart of EEG, indicates that the N1 ERP component originates in the STG (Krumbholz, Patterson, Seither-Preisler, Lammertmann, & Lütkenhöner, 2003). This finding

could reflect the failure of an efference copy to successfully dampen activity in sensory areas after self-produced speech, perhaps due to a delayed corollary discharge (Whitford, et al., 2010). This is supported by neuroimaging evidence suggesting that left superior temporal areas are more active during inner speech in those diagnosed with schizophrenia than healthy controls (Simons et al., 2010). Further EEG studies have shown that theta and gamma band coherence between frontal and temporal areas is impaired in patients with schizophrenia, implying that synchronous neural activity may be the neural substrate of the efference copy (Ford & Mathalon, 2005). Disrupted connectivity between frontal and temporal cortical areas has often been implicated in AVHs (Lawrie et al., 2002), and is possibly linked to structural alterations in white matter tracts such as the arcuate fasciculus (de Weijer et al., 2011). Self-monitoring studies have therefore provided evidence that hallucinating patients may experience inner speech as alien because of a failure of the efference copy system to dampen activity in auditory cortex and label it as self-generated.

It has, however, been argued that a deficit in self-monitoring as measured by some of the aforementioned tasks is not enough in itself to explain the misattribution to *external* sources that has been proposed to explain AVHs, and that there must be a specific bias towards labelling events as external (Allen et al., 2004). Therefore, tasks that attempt to measure participants' bias towards locating events externally have been used with both hallucinating and non-hallucinating patients, as well as healthy controls. These reality monitoring tasks have been used to show a so-called 'externalising bias'. A response bias such as this would lead to a higher likelihood of stimuli of ambiguous source being attributed to an external source (Bentall, 1990). Early tasks utilized signal detection theory, in which hallucination-prone individuals,

hallucinating patients and healthy controls were asked to discriminate whether a voice was present in white noise, showed that the former two groups showed a response bias towards external misattributions (Bentall & Slade, 1985). More recent neuroimaging studies using auditory signal detection tasks have implicated, among other areas, the STG in the creation of false alarms (responding 'yes' when there is no voice present) (Barkus, Stirling, Hopkins, McKie, & Lewis, 2007), therefore showing overlapping regions of activation with neuroimaging studies of both inner speech and AVHs.

A large body of research relating to reality monitoring has also accumulated looking at 'source memory' in people that experience AVHs. In these tasks, participants are generally required to distinguish between self-generated words, experimenter-generated words, and words that have not appeared in the task before (see Waters, Woodward, Allen, Aleman & Sommer, 2012, for a recent review of self-recognition deficits). Findings typically indicate that hallucinating patients or hallucination-prone participants are more likely to misattribute recalled items to the experimenter (Bentall, Baker, & Havers, 1991; Laroi, Van der Linden, & Marczewski, 2004), which has again been taken as evidence that AVHs are linked to an externalising bias. It has also been shown that patients diagnosed with schizophrenia who experience AVHs are more likely to recall an imagined word as spoken (Franck et al., 2000) or an imagined action as performed (Gawęda, Woodward, Moritz, & Kokoszka, 2013), compared to other patients and healthy controls.

Distorted voice tasks have also been used to provide evidence for the existence of an externalising bias in those that experience AVHs. Allen et al. (2004) used a task in which, unlike the aforementioned verbal self-monitoring studies, the speech was pre-recorded. The rationale underlying this alteration was that the task would no longer

measure immediate verbal self-monitoring ability, as participants were not generating the stimuli 'online'. It was found that hallucinating patients were still more likely to make external misattributions. The authors argued that previous findings may not simply be due to a disrupted verbal self-monitoring system, but at least partly due to an externalising bias possibly due to disrupted top-down processing of auditory stimuli. A later neuroimaging study with the same paradigm showed that, in healthy controls and non-hallucinating patients with schizophrenia, the left superior temporal gyrus was generally active when other-produced speech was listened to whereas this was not the case when self-produced speech was listened to. These findings, however, did not apply to hallucinating patients, who did not show differential activity in this area between hearing their own or another's voice (Allen et al., 2007).

The tendency to make external misattributions may therefore be linked to additional, or alternative, mechanisms to the forward model system, because they were gained when participants were not engaged in any motor activity. Allen et al. (2007) suggest that this reflects conscious evaluation of the stimuli, perhaps involving the anterior cingulate cortex (ACC), which has strong connectivity with the temporal cortex (Petrides & Pandya, 1988). Mechelli et al. (2007) have supported this hypothesis by demonstrating a lack of effective connectivity in patients with AVHs, between STG and ACC during other-produced speech. In addition, Vercammen, Knegtering, den Boer, Liemburg & Aleman (2010) have demonstrated atypical functional connectivity of the ACC with left TPJ in hallucinating patients, suggesting that this connectivity may be related to a 'core control network' which exhibits conscious control over experiences.

Table 1.1: Summary of cognitive tasks associated with self-monitoring and reality monitoring, and their association to AVHs

Task	Description	Key findings	Key references
Error monitoring	Participants are asked to monitor their own actions whilst moving a joystick. The proportion of errors corrected is the variable of interest, on the basis that an internal monitor is needed to correct errors made without feedback.	Patients diagnosed with schizophrenia correct errors less often. (Not specific to AVHs.)	Frith and Done (1989)
Distorted voice	Participants listen to recordings of their own voice, and another person's voice. These recordings are sometimes distorted in pitch, and participants respond as to whether they think the voice belongs to them, or not.	Patients with AVHs are more likely to incorrectly respond that a voice belongs to someone else. This finding holds whether the voice is instantly fed back whilst the participant talks, or if it is played back at a later point in time.	Allen et al. (2004) Johns et al. (2001) McGuire, Silbersweig and Frith (1996)
Self-experimenter word production (memory)	Participants must recall whether a word was said by themselves or the experimenter. This task is 'offline', in that it tests performance through *memory* of how an action was performed.	Patients with AVHs are more likely to incorrectly attribute words as produced by the experimenter.	Bentall, Baker and Havers (1991) Laroi, Van der Linden and Marczewski (2004)
Say-imagine word production (memory)	Participants must recall whether they said a word out loud, or imagined it. Alternatively, they may be asked to perform an action, or imagine performing it. This task is 'offline', in that it tests performance through *memory* of how an action was performed.	Patients with AVHs are more likely to incorrectly recall saying a word out loud, or recall performing an action, as opposed to imagining it.	Franck et al. (2000) Gawęda, Woodward, Moritz and Kokoszka (2013)
White noise signal detection (SDT)	Participants listen to bursts of white noise, and must respond, using a button press, whether they think a voice is present in the noise.	Patients with AVHs, and hallucination-prone individuals, make more 'false alarm' responses (hearing voices in white noise that are not present). This seems to be due to a response bias, as opposed to a change in perceptual sensitivity.	Barkus et al. (2007, 2011) Bentall and Slade (1985).

Furthermore, the left planum temporale, an area within the left STG, has been shown to be involved specifically in the perception of externally located speech (Hunter et al., 2002), and posterior parts of the left STG are known to be involved in the spatial localisation of speech (Mathiak et al., 2007); this area has also recently been implicated in the experience of externally as opposed to internally experienced AVHs (Looijestijn et al., 2013). Interestingly, Mathalon, Sullivan, Lim and Pfefferbaum (2001) have shown that the STG diminishes in size over time in those with a diagnosis of schizophrenia, specifically related to positive symptoms in schizophrenia, although they do not report data relating to AVHs, and so it is not possible to tell whether this finding may be specific to AVHs. It has, however, been shown that over time AVHs are more likely to be experienced as internally located (Nayani & David, 1996), although evidence is so far lacking as to whether a correlation exists between this change in STG volume and the likelihood of experiencing AVHs as internal. This evidence, though, implicates left temporal language areas as important in labelling a percept as externally located, and it follows that over-activation of this area may therefore increase the likelihood that a percept will be incorrectly labelled as external.

Temporal lobe regions, then, as well as being important in inner speech and often active during AVHs, have been implicated in both self-monitoring failures in tests of forward model theories and reality monitoring biases towards the external. This may imply that self-monitoring and reality monitoring tasks are to some extent measuring the same cognitive mechanism, although whether this is the ability to distinguish between the internal and external in space, or the ability to monitor self-generated actions and label them as self or non-self, is unknown, and is beyond the scope of this article. Returning to the discussion of neurostimulation as a treatment for AVHs:

neuroimaging findings relating to AVHs, inner speech, and self-/reality monitoring all point towards a key role for the left temporal lobe in the experience of AVHs (with differential findings regarding the right hemisphere). It is therefore possible that the success of the treatment may depend on its ability to modulate cortical areas involved in inner speech and self-/reality monitoring.

4. Neurostimulation as a treatment for AVHs

4.1. Neurostimulation as a therapeutic technique

Transcranial magnetic stimulation (TMS) is a noninvasive brain stimulation technique in which a coil placed on the scalp uses a rapidly changing magnetic field to induce an electrical current in the cortex (Hallett, 2007; Walsh & Cowey, 2000). Pioneered by Barker, Freeston, Jalinous, Merton, and Morton (1985), TMS was at first used in single pulses, and can essentially introduce a focal area of neural noise in an area of cortex by activating neurons underlying the stimulating coil. Repetitive TMS (rTMS), in contrast, uses repeated pulses and can be applied in an event-related manner (to disrupt regions, synchronously with presented stimuli), to test whether a specific cortical area is necessary when completing a specific cognitive task (as, if the area is responding to the magnetic pulses, it cannot respond to the concurrent task demands). It is worth noting that secondary areas may also be affected by the introduction of neural noise due to connectivity, and that task effects may not be related to rTMS of the primary region but functionally connected, or indeed anatomically connected, regions (Komssi et al., 2002; Walsh & Pascual-Leone, 2003). This factor is most prevalent when rTMS is utilized over longer time periods, in the absence of any task or stimuli, because it can have lasting after-effects of

excitation or inhibition of cortical areas both directly underneath the coil and trans-synaptically (Hoffman & Cavus, 2002; Wassermann, Wedegaertner, Ziemann, George, & Chen, 1998). Results showing changes in excitation or inhibition in regions distal to the stimulating coil highlight that inferences regarding the role of specific brain areas in tasks need to be made cautiously, although this may be an advantage when attempting to modulate activity in widespread cortical networks (Pascual-Leone et al., 1998).

Although TMS excites all neurons in the stimulated region with each pulse (both excitatory and inhibitory), it is important to distinguish between this and the excitation or inhibition of function that may follow. For example, low frequency (1 Hz) rTMS can have lasting after-effects, which tends to cause a decrease in neuronal activity in the stimulated region, whereas higher frequencies (> 5 Hz) can cause lasting excitation (Maeda, Keenan, Tormos, Topka, & Pascual-Leone, 2000). The effects of TMS can also be modulated by underlying tissue type. For example, differences in anisotropy can affect the spatial distribution of the induced field, so although TMS is typically thought to largely affect grey matter, recent findings indicate that the morphology of underlying white matter tracts is also important (De Lucia, Parker, Embleton, Newton, & Walsh, 2007; Opitz et al., 2013). It has further been established that TMS can have state-dependent effects, and is thought to preferentially stimulate neurons that are less active (Silvanto, Muggleton, & Walsh, 2008; Silvanto, Muggleton, Cowey, & Walsh, 2007) and can also have differential effects on cortical excitability depending on baseline levels. For example, Siebner et al. (2004) showed that if excitability was increased at baseline, then 1 Hz rTMS reduced excitability; however, if excitability was decreased at baseline, the same stimulation had the effect of increasing excitability. The mechanism through which

rTMS can produce lasting after-effects is still somewhat unclear, but may be due to long-term potentiation or long-term depression (LTP/LTD)-like effects, i.e., the observation that the strength of synapses between neurons can be altered if they repeatedly fire synchronously (Hoffman & Cavus, 2002). Since rTMS can have effects in regions distal to the stimulating coil, particularly when used to produce after-effects, it has the potential to affect neuronal networks thought to be involved in neurological and psychiatric conditions that may be a result of changes in connectivity between brain regions.

Typically, studies applying rTMS to test its therapeutic potential stimulate for protracted periods of time (e.g., 15 minutes per day, for three weeks). The intensity of stimulation is determined for each participant separately, using the individual's 'motor threshold' (the intensity at which stimulation of motor cortex can elicit a hand movement); for example, treatment may be administered at 90% of each individual's motor threshold. Perhaps most famously, the observation that under-activation of the left dorsolateral prefrontal cortex often coincides with clinical depression led to the use of high-frequency rTMS over this area, and there is evidence that it may be an effective treatment option (George et al., 2010; George et al., 1995). That said, some argue that the efficacy has often been exaggerated, and more studies may be needed to ensure that improvements are not simply a placebo effect (Miniussi et al., 2005; Ridding & Rothwell, 2007). The example of depression highlights difficulties in showing efficacy of rTMS as a valid treatment option, in that depression is a diverse diagnosis, and by definition is a subjective experience. This makes it hard to exclude placebo effects, especially since the control 'sham' condition usually used in rTMS studies has been criticized (Robertson, Théoret, & Pascual-Leone, 2003). These criticisms are equally valid when applied to using

rTMS to treat AVHs, and will be returned to below. Nevertheless, rTMS has now been approved for use in the treatment of depression by the Food and Drug Administration in the US (Connolly, Helmer, Cristancho, Cristancho, & O'Reardon, 2012).

An alternative neurostimulation technique is transcranial direct current stimulation (tDCS). This can be used to selectively increase or decrease excitability of brain areas, as rTMS can. In tDCS, a weak electrical current is passed between two electrodes attached to the scalp. Current runs from an anodal electrode, under which the neurons' membrane potentials are generally depolarized, to the cathodal electrode, under which they are generally hyperpolarized. This leads to increased neuronal excitability under the anode, and decreased excitability under the cathode (Nitsche & Paulus, 2000, 2011). Importantly, effects of tDCS which outlast the stimulation period are often observed, probably due to longer term GABAergic and glutamatergic mechanisms (Stagg & Nitsche, 2011), leading to studies into whether this technique could be used therapeutically for neurological and psychiatric disorders. One advantage of using tDCS over rTMS in an experimental setting is a more realistic sham condition. Active stimulation using tDCS leads to no more than a tingling or itching sensation underneath the electrodes, and participants tend to report that this sensation fades away after a short period of time. Therefore, sham tDCS attempts to mimic this by stimulating for only 30 seconds, and then gradually decreasing the stimulation intensity until the equipment is turned off. In this way, participants tend to be unaware that they are no longer receiving active stimulation (Gandiga, Hummel & Cohen, 2006; though see O'Connell et al., 2012). On top of this, tDCS is less expensive and easier to apply than rTMS, and can potentially be

used by patients at their own homes, with the clinician providing indirect support with a remote trigger (Brunoni et al., 2012).

4.2. Can neurostimulation be used to treat AVHs?

As discussed above, neuroimaging studies using positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) have shown that AVHs are often accompanied by activation of the speech and language perception areas in the left hemisphere, in agreement with inner speech theories of AVHs (Allen et al., 2012; Silbersweig et al., 1995), and research also suggests that patients with AVHs often show deficits in speech processing (Hoffman, Rapaport, Mazure, & Quinlan, 1999). Therefore, initial studies tested the therapeutic effect of low-frequency (1 Hz) rTMS over left temporoparietal cortex (midway between the T3 and P3 electrodes using the EEG 10-20 system), at first tested on three patients diagnosed with schizophrenia (Hoffman et al., 1999), and later on a larger sample of 50 (Hoffman et al., 2005). Hallucinating patients received rTMS treatment for 9 consecutive days (excluding weekends). These initial studies indicated that rTMS may be effective as a treatment to reduce AVHs, as measured by the Auditory Hallucinations Rating Scale, a 7-item scale which assesses hallucination frequency, number of voices, volume, vividness, salience, length and distress caused. A large effect size of .94 was found in the 50 patient sample, reducing the *frequency* of AVHs. There was no improvement in other scores relating to positive or negative symptoms of schizophrenia, implying that the effects of stimulation are relatively specific to a reduction in AVHs.

Recently, Hoffman et al. (2013) replicated their initial findings with a sample size of 83, albeit it with reduced effect size of .65 for reduction in frequency of AVHs. This

35

effect size was increased to .74 when only patients with whom they could consistently detect a motor threshold were included. They also showed that stimulation of the right homologue of Wernicke's area could lead to a reduction in frequency of AVHs, especially for those rated high in 'attentional salience' ("the degree to which hallucinations capture attention and alter ongoing thoughts and behaviour", p. 2).

Some studies, however, have failed to show the substantial improvement reported by Hoffman and colleagues. Notably, a relatively large randomized controlled trial (N = 62) failed to find any significant advantage of active rTMS over sham stimulation, despite using fMRI and image-guided stereotaxy to localize the stimulation to the point of maximal activity during each patient's AVHs (Slotema et al., 2011). Nevertheless, meta-analyses with this finding taken into account still showed positive effects of rTMS with a moderate effect size of .38 (Slotema & Daskalakis, 2012).

Whether rTMS is effective at reducing frequency of hallucinations is confounded by the fact that most studies have used either medication- or therapy-resistant patients with a diagnosis of schizophrenia. It is difficult to speculate on whether rTMS would be more or less effective if tested on drug-naive individuals, or on patients without this diagnosis, but inter-individual variability in, for example, white matter volume, could change the distribution of current induced by stimulation, as previously mentioned. Many studies also do not report the specificity of the effects of neurostimulation – that is, whether there was a corresponding reduction in other positive symptoms. This information is crucial if conclusions are to be drawn relating to the underlying mechanisms of AVHs. Nevertheless, in Hoffman et al.'s (2005) study, it is reported that there was no significant change in the frequency of

other positive or negative symptoms, and so it seems likely that neurostimulation treatment operates solely on the symptom of AVHs.

Additionally, it is possible that a publication bias has meant that negative findings with regard to efficacy are not publicly available; the effect sizes of published studies using rTMS to treat AVHs have tended to decrease with time, and so it is possible that some early negative findings were not published. Nevertheless, Slotema, Aleman, Daskalakis and Sommer (2012) conclude that there is little evidence of a publication bias, because there are examples in the literature of small, early studies with negative findings. Another major criticism, also made by Slotema et al., is that many studies have not achieved adequate statistical power: Hoffman et al.'s 50 patient sample, their recent 83 patient sample, and Slotema et al.'s recent negative finding being the exceptions. It is important that future studies aim to achieve higher statistical power to increase reliability.

One possible reason for the variable findings of the therapeutic effects of rTMS on AVHs is the inadequacy of the sham condition in rTMS trials. Active rTMS trials elicit a loud clicking sound, with a characteristic tapping sensation on the scalp underneath the coil. This sensation is hard to mimic realistically – although sham coils do exist, they usually do not mimic anything more than the auditory aspect of receiving rTMS. Many studies, rather than utilising a sham coil, will tilt the active rTMS coil 45° or 90° from the scalp. In this way, both the sound and tactile sensation of rTMS are, to some extent, replicated. This method, though, leaves open questions about whether the stimulation may have some effect on underlying (or surrounding areas of) cortex. Indeed, Lisanby, Gutman, Luber, Schroeder, and Sackeim (2001) showed that, when the coil was tilted 45° from the scalp, the voltage induced in the cortex was approximately 33% of that induced in the 'active TMS'

condition. It is also unclear to what extent this technique is successful in blinding participants to the condition they are in: Hoffman et al. (2005), using this sham technique, reported that many patients correctly guessed which condition they had been in, but argued that in most cases their guess was actually based on curtailment of symptoms. Nevertheless, this is anecdotal, and the fact that many patients were able to tell which condition they were in may have affected the results.

It has therefore been suggested that in experiments showing a positive effect, patients receiving rTMS treatment in fact showed a placebo effect to the treatment, with the observed difference between conditions being due to an inadequate sham condition (Slotema, et al., 2011). Further, the essentially subjective measures of severity of hallucinations arguably leave the studies even more susceptible to being confounded by the placebo effect. Some measures, however, are less subjective than others; Hoffman et al. (2013) asked participants to record frequency of AVHs with a mechanical counter, perhaps negating this criticism. In addition, most rTMS studies have recruited patients whose AVHs have been refractory to anti-psychotic drugs or other treatment options, leaving open the question of why the placebo effect would be evident after rTMS, but not other attempts to reduce AVHs.

No studies to date have looked at the efficacy of rTMS in the treatment of different types of AVH, but there is a growing realisation that AVHs cannot be treated as one homogeneous group, and may in fact differ both phenomenologically and in their cognitive and neural substrates (Jones, 2010; Nayani & David, 1996). Recent studies, for example, have suggested a subtype of AVH known as 'hypervigilance hallucinations', characterised by their occurrence when attention is externally focused (Garwood, Dodgson, Bruce, & McCarthy-Jones, 2013). That phenomenologically different AVHs may have different neural substrates is

highlighted by Hoffman et al.'s most recent study (2013), showing that the effectiveness of rTMS to the right hemisphere is dependent on attentional salience of the AVHs. It is therefore possible that only some AVHs may be amenable to treatment using rTMS. If this were the case, the results may be skewed depending on the 'types' of AVHs that were studied. It will be argued in Section 4.3 that AVHs which may be best described as misattributed inner speech may be most amenable to neurostimulation treatment.

At the time of writing, only one experimental study of the therapeutic effect of tDCS on AVHs has been reported. Brunelin et al. (2012) studied 30 hallucinating individuals diagnosed with schizophrenia, placing the cathodal electrode over the left temporoparietal junction (midway between the T3-P3 electrodes as specified by the 10-20 EEG system, similarly to rTMS studies), and the anodal electrode over left dorsolateral prefrontal cortex; abnormal white matter volume of this area is often associated with negative symptoms of schizophrenia (Sanfilipo et al., 2000). The participants underwent stimulation twice a day, for 20 minutes, for 5 days. Half of the participants were assigned to the sham condition, while half were assigned to receive active stimulation. Results showed that those who received active cathodal stimulation over temporoparietal cortex experienced a 31% reduction in hallucination severity, as measured by the Auditory Hallucination Rating Scale (which takes into account variables such as hallucination frequency, loudness, and salience of the AVH), compared to an 8% reported reduction in the sham condition. This effect was still present 3 months later, with 6 of 15 participants in the experimental condition showing a reduction in hallucination frequency of more than 50%. The most obvious criticism of this study is the relatively small sample size; it will be interesting to see whether future studies are able to replicate these results. As

already mentioned, the sham condition used in studies using tDCS is more effective, and therefore may be less susceptible to placebo effects. Issues such as portability and ease of use may also make it more realistic as a treatment option.

To summarize, the evidence regarding efficacy of noninvasive brain stimulation techniques as a treatment for AVHs is still equivocal. What follows is an analysis of how treatment of AVHs with neurostimulation may affect the associated cognitive and neural mechanisms, interpreted within an inner speech monitoring framework.

4.3. How might treatment with neurostimulation affect the cognitive and neural mechanisms associated with AVHs?

Previous attempts to treat AVHs with noninvasive brain stimulation have not been carried out based on a clear prediction from our understanding of the cognitive mechanisms underlying AVHs. Recent evidence, however, suggests that neurostimulation may be effective as a treatment option due to its effects on brain networks involved in the monitoring of inner speech. The following section aims to interpret the findings discussed in Section 4, based on inner speech models of AVHs, and discuss recent studies which have used neuroimaging to monitor the after-effects of neurostimulation in cortical regions known to be involved in AVHs. Table 1.2 provides a summary of some key findings regarding the importance of different brain areas in AVHs and neurostimulation treatment.

One of the few studies to look at the effects of left temporoparietal rTMS on reality monitoring performance in those that experience AVHs was conducted by Brunelin et al. (2006). This study used an rTMS protocol similar to those used in other studies of rTMS to treat AVHs, and was able to replicate the improvements in Auditory Hallucination Rating Scale score shown by others. The 24 patients also took part in a

source memory test in which they had to recall whether they had read an item silently to themselves, or said the word out loud (a 'say/imagine' paradigm). Patients that received active rTMS were less likely to misattribute an imagined word as one they had said after stimulation, whereas those that were allocated to the sham condition did not show this pattern. This can therefore be seen to support both the efficacy of rTMS as a treatment option for AVHs, and the link between AVHs and reality monitoring.

The importance of the STG in responders to neurostimulation treatment has also been highlighted by a recent neuroimaging study showing higher left STG activation in a pre-TMS resting-state scan in those that were later classified as 'responders' to the TMS treatment paradigm (Homan, Kindler, Hauf, Hubl, & Dierks, 2012). These findings may be crucial, as they suggest that pre-existing levels of activity in the STG may be one biomarker for recognizing likely responders to neurostimulation treatment, and are also consistent with findings discussed in Section 4.1 showing that the after-effects of rTMS are dependent on baseline levels of excitability. In a separate study, measurement of cerebral blood flow post-TMS treatment, relative to pre-treatment, showed reductions in activation in primary auditory cortex (part of the STG), Broca's area and the cingulate gyrus after 10 days of rTMS treatment (Kindler et al., 2013). This provides support for the claim that left temporoparietal cortex (the site of stimulation) is the most appropriate location to affect other temporal regions, as well as cingulate areas that may be related to conscious evaluation of stimuli. Kindler et al. suggest that a high level of activity in the STG makes it harder to differentiate between inner speech and externally perceived speech, concordant with self- or reality monitoring cognitive models of AVHs. This evidence is also consistent, to some extent, with some forward model theories of AVHs which argue

that a failure of the efference copy mechanism leads to an over-active auditory cortex, therefore leading to self-produced speech being labelled as 'non-self' (although it neither supports nor contradicts ideas surrounding the cause of over-activity). Importantly, only reduced activation in the STG correlated with a reduction in AVHs, implying that effects in other areas may be not be causal to the improvement. The authors argue that the activation of Broca's area typically seen during AVHs is due to the production of (subsequently misattributed) inner speech, although whether the observed reduction in activity post-TMS treatment corresponds to any change in the experience of inner speech (phenomenological or quantitative) is unknown.

An excellent addition to these studies, in our opinion, would have been to investigate if any phenomenological differences in AVHs existed between the responders and non-responders to rTMS treatment. For example, it is possible that only AVHs that are best categorized as 'misattributed inner speech' may be amenable to this form of treatment. This hypothesis is consistent with imaging studies of self- and reality monitoring studies that show that activation of this area is related to task performance, as well as activation seen when participants are engaged in inner speech, as outlined in Sections 2 and 3. Hoffman et al.'s finding (2013) of efficacy of right-sided stimulation in more salient AVHs further implies that phenomenologically different AVHs may need to be accounted for by different cognitive-neuroscientific models. Alternatively, it is possible that a higher level of activation in right-sided temporal areas leads to a higher level of attentional salience because of other phenomenological variables, such as emotional valence.

Results from rTMS treatment protocols, however, have also been used to argue that cortical areas involved in self-agency (as measured by self-monitoring and reality

monitoring tasks) can in fact be dissociated from those involved in AVHs. Jardri et

al. (2009) showed that low-frequency stimulation of the right TPJ could improve

performance on both types of task, but not decrease frequency of AVHs; meanwhile,

left TPJ stimulation achieved both. The data, though, was gained from only a single

participant (a child diagnosed with childhood-onset schizophrenia), and so further

research is needed to support this finding. Despite this, the authors suggest that self-

agency may be linked to a network dissociable to that drawn upon during AVHs,

implicating the inferior parietal lobule (IPL) rather than the TPJ/STG per se.

Previous studies have suggested bilateral involvement of the IPL in feelings of

agency, with a right sided dominance (Farrer et al., 2003; Jardri, et al., 2007). The

IPL is immediately adjacent to the posterior STG (including the planum temporale

and Wernicke's area), and may be part of an alternative pathway that runs laterally to

the arcuate fasciculus between Wernicke's and Broca's area (Catani, Jones, &

ffytche, 2005; Frey, Campbell, Pike, & Petrides, 2008). It is possible that the IPL is

part of a feed-forward mechanism between speech production areas and temporal

areas (Rauschecker & Scott, 2009), which may explain evidence of its involvement

in feelings of agency and monitoring tasks. It is therefore difficult to ascertain

whether concurrent improvement in monitoring tasks and AVH frequency is due to

the effect of stimulation on the targeted area (left TPJ) and its connections with other

auditory cortical areas, or the immediately adjacent IPL, due to either the limitations

of the spatial resolution of rTMS (approximately 1cm^3, dependent on coil size), or

connectivity between the STG and the left IPL.

Table 1.2: Brain areas important in understanding the effects of neurostimulation on AVHs, and their connectivity with other regions.

Brain region	Role in AVHs	Relevance to neurostimulation treatment	Connectivity
Superior temporal gyrus (STG)	Includes PAC, Wernicke's area, and planum temporale. Structural abnormalities and functional activity consistently implicated in AVHs, and during monitoring tasks.	Posterior STG activity reduced after neurostimulation; this correlates with reduction in AVH severity.	Strong connectivity with TPJ, and effective connectivity with ACC. Also connected to IFG through arcuate fasciculus white matter tract.
Inferior frontal gyrus (IFG, Broca's area)	Crucial for production of speech (including inner speech), particularly in the left hemisphere. Role of right IFG still relatively unexplored.	rTMS of Broca's area does not lead to a reduction in AVH frequency. Reduction in activity in IFG following stimulation of left TPJ, though not correlated with reduction in AVH frequency.	Connected to STG through arcuate fasciculus white matter tract. Excessive functional connectivity with putamen in voice-hearers.
Anterior cingulate cortex (ACC)	Activation seen during AVHs may reflect conscious evaluation of stimuli, and in combination with STG, may be involved in monitoring processes.	Reduction in activity in ACC following stimulation of left TPJ, though not correlated with reduction in AVH frequency.	Connectivity with STG & TPJ may reflect verbal monitoring processes – effective connectivity during monitoring task is reduced in voice-hearers.
Inferior parietal lobe (IPL)	Often activated in symptom-capture studies of AVHs, and commonly linked to feelings of self-agency.	Data from neuroimaging has not implicated changes in activation post-neurostimulation; however, close proximity could mean activity is modulated by TPJ stimulation.	May be part of an alternative pathway that runs laterally to the arcuate fasciculus, between IFG and STG.
Putamen	Hoffman's corticostriatal loop model specifies that an overabundance of language representations initiated by the putamen may surface 'unbidden thoughts' as AVHs, due to hyperconnectivity with STG and IFG.	If Hoffman's model is supported, disruption of hyperconnectivity with this region may be related to the therapeutic effect of neurostimulation.	Excessive functional connectivity with IFG and STG in voice-hearers.

STG = superior temporal gyrus; PAC = primary auditory cortex; TPJ = temporoparietal junction; ACC = anterior cingulate cortex; IFG = inferior frontal gyrus; IPL = inferior parietal lobe.

This is difficult to reconcile with imaging studies of monitoring tasks and symptom-capture studies of AVHs, in which superior temporal regions, particularly on the left, are often shown to be important. Although inferior parietal regions have been reported as involved in the occurrence of AVHs (Jardri, Pouchet, Pins, & Thomas, 2011; Lennox, Park, Medley, Morris, & Jones, 2000), one might expect more reliable activation of this area in neuroimaging studies of AVHs if it were of such key importance. Speculations regarding the importance of the IPL in monitoring and AVHs therefore need to be empirically tested with larger sample sizes and a variety of neuroscientific and cognitive tests, before firm conclusions can be drawn.

4.4. Neurostimulation and inner speech

It is interesting that Hoffman et al. (2007) found no reduction in AVHs after stimulating Broca's area, but did show that responders to the rTMS treatment tended to show reduced functional connectivity between left TPJ and the right homologue of Broca's area, supporting Vercammen et al. (2010), who found reduced connectivity between these areas in patients with AVHs. In combination with Aziz-Zadeh et al.'s (2005) results showing induction of covert speech arrest during Broca's area stimulation, the lack of improvement shown after treatment through Broca's area stimulation may support arguments that there is no difference, at a neural level, in the *production* of inner speech in those that experience AVHs. Instead, it is the subsequent perception (by temporal regions) or evaluation (by ACC) that leads to a misattribution. In combination with the aforementioned study by Homan et al. (2012), these results indicate that reduction of activity in superior temporal regions is crucial to the therapeutic effect of rTMS on AVHs. More detailed studies are required to distinguish whether reduction in activity in specific areas of the STG,

such as the planum temporale, are responsible for the improvement in AVH frequency.

The role of right hemispheric language areas in inner speech and AVHs is still a relatively unexplored area: current inner speech theories might predict that they are not integral to the creation of AVHs, but are instead important in their contextualisation (experiencing AVHs as another person's voice and/or perspective, emotional content of AVHs). Nevertheless, neurostimulation treatment protocols reviewed above suggest that in some cases stimulation of the right TPJ may be successful in reduction of AVH *frequency*, which may imply a causal role for right temporal regions in the experience of AVHs, rather than simply contextualisation. Further research into the neural basis of different forms of inner speech (e.g., dialogic, emotional) could help to clarify what role the right homologue of Broca's area and right superior temporal regions play in the generation of inner speech, self-/source monitoring, and therefore AVHs.

5. Future directions

The evidence so far reviewed suggests that the efficacy of neurostimulation as a treatment option may depend on its ability to modulate activity in superior temporal cortical regions, as well as inferior frontal and anterior cingulate regions. We have argued that this is consistent with inner speech theories of AVHs, which postulate that atypical monitoring processes lead to its misattribution to an external or non-self source. This is supported by findings implicating similar regions in the monitoring of speech as those affected by neurostimulation of TPJ. There are, however, a number of key avenues of research which remain to be explored.

While it seems likely that, at a cognitive level, the effect of treatment may be due to an improvement in self- and/or reality monitoring, it is hard to discount the possibility that a decrease in AVH frequency may lead to improvements in monitoring capability, perhaps due to the distracting effects of AVHs. Future studies should aim to test this more directly. Such studies could use low frequency rTMS, or tDCS, in conjunction with previously used monitoring tasks in both clinical and non-clinical populations, to study the effects of reduction of STG activity on the ability to distinguish between internal and external events. Such studies would not only be able to test the importance of superior temporal regions in monitoring tasks, but would also be able to inform us of the causality of observed improvements. They would also potentially provide evidence both for cognitive models that specify monitoring deficits or biases as a cause of AVHs, and clinical studies of the efficacy of brain stimulation techniques that claim to be able to reduce AVH frequency.

As already noted, it is likely that AVHs are not a homogeneous phenomenon (Jones, 2010). The experience may differ, for example, in terms of level of externality (the extent to which the voice is experienced as coming from the external environment), the number of voices heard, the volume of the voice, and attentional salience (the extent to which the voice captures the attention of the person, and is thus effective in altering behavior). As yet, only a small number of studies have investigated differences in neural activity between phenomenologically different AVHs using fMRI (Looijestijn et al., 2013; Vercammen, Knegtering, Bruggeman, & Aleman, 2010). Hoffman et al. (2013) has noted that attentional salience appears to be one marker for likely response to neurostimulation treatment to the right TPJ (as opposed to the left TPJ).

On a broader level, it is possible that some AVHs can best be described as misattributed inner speech, whereas others might best be described as intrusive memories (Waters, Badcock, Michie, & Maybery, 2006), and others still as 'hypervigilance' towards detecting stimuli in the environment (Dodgson & Gordon, 2009; Garwood, Dodgson, Bruce, & McCarthy-Jones, 2013). This review has focused on the former; that is, AVHs which seem to be explicable within an inner speech framework. As outlined, it may be that neurostimulation of the TPJ affects mechanisms that are involved with the monitoring of inner speech. An important area for future research would therefore be to investigate whether these subtypes of hallucinations are distinguishable by neural activation, and whether some types of AVHs are more amenable to treatment with neurostimulation.

Thirdly, there is a need to develop more valid inner speech paradigms if we are to understand its relation to AVHs. For example, ideas surrounding the dialogic nature of inner speech are yet to be tested within a cognitive-neuroscientific framework – important questions to address would be related to differences in activation between monologic inner speech (inner speech that does not involve the back-and-forth of a conversation) and dialogic inner speech. Moreover, it is important that more *realistic* forms of inner speech are studied. Currently, most studies rely on asking participants to repeat sentences to themselves (e.g., McGuire et al., 1995) or requiring participants to count syllables (e.g., Aziz-Zadeh et al., 2005). More valid forms of inner speech could be evoked by, for example, asking participants to imagine a conversation, evaluate behaviour, or plan a speech for a future event – these have all been suggested as important functions of inner speech (McCarthy-Jones & Fernyhough, 2011), and so should be more accurate approximations. A further possibility would be to use experience sampling techniques such as Descriptive

Experience Sampling, which involves fitting participants with a 'beeper' which randomly cues the participant to report their current inner experience (Hurlburt & Heavey, 2001). Coupled with neuroimaging techniques, this could become a powerful method by which to investigate the neural mechanisms underlying inner speech and AVHs.

Importantly, there is a need for larger scale tests of treating AVHs with neurostimulation. Of the three studies discussed here that have achieved adequate power, one finds no effect of neurostimulation (Slotema et al., 2011), and the other two originate from the same institution (Hoffman et al., 2005; 2013). Further replication studies by independent teams are needed, and tests should be consistent with the methods used to target the stimulation (e.g., consistently using structural or functional MRI scans to locate TPJ, or following the EEG 10-20 system). It is important that studies attempt to monitor possible effects of neurostimulation of left TPJ on a wide variety of variables. Currently, evidence suggests that there are no negative effects on neuropsychological measures such a short-term verbal memory (Hoffman et al., 2005) or measures of hearing function such as pure-tone audiometry (Schonfeldt-Lecuona et al., 2012), but it would also be interesting to study potential changes in phenomenal characteristics of inner speech following reduction in activity in either left or right superior temporal regions. Finally, as outlined, testing the efficacy of tDCS to treat AVHs is a promising area of research and studies with larger samples are needed to examine whether this technique could be a useful addition to currently available treatment options for those who seek help in relieving the distress of AVHs.

6. Conclusions

Using neurostimulation as a treatment option for AVHs seems promising. Existing findings indicate that over-activation of the STG in the resting state is one marker for a response to the treatment. If it is possible, finding phenomenological markers of likely responders would not only mean that treatment could be targeted quickly and easily to those who might benefit most, but would also tell us much about the underlying cognitive neuroscience of AVHs. Although controversy still exists as to whether the putative therapeutic effects of rTMS can simply be attributed to an ineffective sham condition, future studies, especially those using tDCS, could settle this debate. Indeed, if noninvasive brain stimulation techniques are to be taken seriously as a viable treatment option, tDCS is a much more realistic alternative, due to the portability, ease and comfort of use, and cost. Further study is also needed into the long-term effects on AVHs of this treatment – currently, minimal evidence exists into the effects past one month.

Models of AVHs that suggest self-monitoring deficits or reality monitoring biases, leading to the misattribution of inner speech to an external or non-self source, do seem to be supported by studies using brain stimulation techniques. Although it is important not to overstate the power of neurostimulation as an experimental technique, neuroimaging studies of both hallucinating individuals and of individuals performing monitoring tasks point to the importance of left superior temporal regions and areas connected functionally and anatomically to it, in these processes. Typical neurostimulation protocols, meanwhile, direct the stimulation to affect these areas. There is tentative evidence that improvement in AVH frequency following rTMS coincides with improvement on monitoring tasks, although much more work needs to be carried out in this area to establish a causal link between the two.

References

Alba-Ferrara, L., Ellison, A., & Mitchell, R. L. C. (2012). Decoding emotional prosody: resolving differences in functional neuroanatomy from fMRI and lesion studies using TMS. *Brain Stimulation, 5*(3), 347-353.

Alba-Ferrara, L., Fernyhough, C., Weis, S., Mitchell, R. L. C., & Hausmann, M. (2012). Contributions of emotional prosody comprehension deficits to the formation of auditory verbal hallucinations in schizophrenia. *Clinical Psychology Review, 32*(4), 244-250.

Allen, P., Amaro, E., Fu, C. H. Y., Williams, S. C. R., Brammer, M. J., Johns, L. C., et al. (2007). Neural correlates of the misattribution of speech in schizophrenia. *British Journal of Psychiatry, 190*, 162-169.

Allen, P., Johns, L. C., Fu, C. H. Y., Broome, M. R., Vythelingum, G. N., & McGuire, P. K. (2004). Misattribution of external speech in patients with hallucinations and delusions. *Schizophrenia Research, 69*(2–3), 277-287.

Allen, P., Larøi, F., McGuire, P. K., & Aleman, A. (2008). The hallucinating brain: a review of structural and functional neuroimaging studies of hallucinations. *Neuroscience & Biobehavioral Reviews, 32*(1), 175-191.

Allen, P., Modinos, G., Hubl, D., Shields, G., Cachia, A., Jardri, R., et al. (2012). Neuroimaging auditory hallucinations in schizophrenia: from neuroanatomy to neurochemistry and beyond. *Schizophrenia Bulletin, 38*(4), 695-703.

Aziz-Zadeh, L., Cattaneo, L., Rochat, M., & Rizzolatti, G. (2005). Covert speech arrest induced by rTMS over both motor and nonmotor left hemisphere frontal sites. *Journal of Cognitive Neuroscience, 17*(6), 928-938.

Barker, A. T., Freeston, I. L., Jalinous, R., Merton, P. A., & Morton, H. B. (1985). Magnetic stimulation of the human brain. *Journal of Physiology (London), 369,* 3P.

Barkus, E., Stirling, J., Hopkins, R., McKie, S., & Lewis, S. (2007). Cognitive and neural processes in non-clinical auditory hallucinations. *British Journal of Psychiatry, 191*, S76-S81.

Beavan, V., Read, J., & Cartwright, C. (2011). The prevalence of voice-hearers in the general population: a literature review. *Journal of Mental Health, 20*(3), 281-292.

Bentall, R. P. (1990). The illusion of reality - a review and integration of psychological research on hallucinations. *Psychological Bulletin, 107*(1), 82-95.

Bentall, R. P. (2003). *Madness Explained*. London: Penguin.

Bentall, R. P., Baker, G. A., & Havers, S. (1991). Reality monitoring and psychotic hallucinations. *British Journal of Psychiatry, 30*, 213-222.

Bentall, R. P., & Slade, P. D. (1985). Reality testing and auditory hallucinations - a signal detection analysis. *British Journal of Psychiatry, 24*, 159-169.

Blakemore, S. J., Smith, J., Steel, R., Johnstone, E. C., & Frith, C. D. (2000). The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: evidence for a breakdown in self-monitoring. *Psychological Medicine, 30*(5), 1131-1139.

Brunelin, J., Mondino, M., Gassab, L., Haesebaert, F., Gaha, L., Suaud-Chagny, M. F., et al. (2012). Examining transcranial direct-current stimulation (tDCS) as a treatment for hallucinations in schizophrenia. *American Journal of Psychiatry, 169*(7), 719-724.

Brunelin, J., Poulet, E., Bediou, B., Kallel, L., Dalery, J., D'Amato, T., et al. (2006). Low frequency repetitive transcranial magnetic stimulation improves source monitoring deficit in hallucinating patients with schizophrenia. *Schizophrenia Research, 81*(1), 41-45.

Brunoni, A. R., Nitsche, M. A., Bolognini, N., Bikson, M., Wagner, T., Merabet, L., et al. (2012). Clinical research with transcranial direct current stimulation (tDCS): challenges and future directions. *Brain Stimulation, 5*(3), 175-195.

Buchanan, R. W., Kreyenbuhl, J., Kelly, D. L., Noel, J. M., Boggs, D. L., Fischer, B. A., et al. (2010). The 2009 Schizophrenia PORT psychopharmacological treatment recommendations and summary statements. *Schizophrenia Bulletin, 36*(1), 71-93.

Carlsson, A. (1978). Anti-psychotic drugs, neurotransmitters, and schizophrenia. *American Journal of Psychiatry, 135*(2), 164-173.

Catani, M., Jones, D. K., & ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Annals of Neurology, 57*(1), 8-16.

Connolly, K. R., Helmer, A., Cristancho, M. A., Cristancho, P., & O'Reardon, J. P. (2012). Effectiveness of transcranial magnetic stimulation in clinical practice post-FDA approval in the United States: results observed with the first 100 consecutive cases of depression at an academic medical center. *Journal of Clinical Psychiatry, 73*(4), e567-573.

De Lucia, M., Parker, G. J. M., Embleton, K., Newton, J. M., & Walsh, V. (2007). Diffusion tensor MRI-based estimation of the influence of brain tissue anisotropy on the effects of transcranial magnetic stimulation. *NeuroImage, 36*(4), 1159-1170.

de Weijer, A. D., Mandl, R. C. W., Diederen, K. M. J., Neggers, S. F. W., Kahn, R. S., Pol, H. E. H., et al. (2011). Microstructural alterations of the arcuate fasciculus in schizophrenia patients with frequent auditory verbal hallucinations. *Schizophrenia Research, 130*(1–3), 68-77.

Dodgson, G., & Gordon, S. (2009). Avoiding false negatives: are some auditory hallucinations an evolved design flaw? *Behavioural and Cognitive Psychotherapy, 37*(3), 325-334.

Farde, L. (1997). Brain imaging of schizophrenia – the dopamine hypothesis. *Schizophrenia Research, 28*(2–3), 157-162.

Farrer, C., Franck, N., Georgieff, N., Frith, C. D., Decety, J., & Jeannerod, M. (2003). Modulating the experience of agency: a positron emission tomography study. *NeuroImage, 18*(2), 324-333.

Fernyhough, C. (1996). The dialogic mind: a dialogic approach to the higher mental functions. *New Ideas in Psychology, 14*(1), 47-62.

Fernyhough, C. (2004). Alien voices and inner dialogue: towards a developmental account of auditory verbal hallucinations. *New Ideas in Psychology, 22*(1), 49-68.

Ford, J. M., & Mathalon, D. H. (2005). Corollary discharge dysfunction in schizophrenia: can it explain auditory hallucinations? *International Journal of Psychophysiology, 58*(2-3), 179-189.

Ford, J. M., Mathalon, D. H., Heinks, T., Kalba, S., Faustman, W. O., & Roth, W. T. (2001). Neurophysiological evidence of corollary discharge dysfunction in schizophrenia. *American Journal of Psychiatry, 158*(12), 2069-2071.

Franck, N., Rouby, P., Daprati, E., Daléry, J., Marie-Cardine, M., & Georgieff, N. (2000). Confusion between silent and overt reading in schizophrenia. *Schizophrenia Research, 41*(2), 357-364.

Frey, S., Campbell, J. S. W., Pike, G. B., & Petrides, M. (2008). Dissociating the human language pathways with high angular resolution diffusion fiber tractography. *Journal of Neuroscience, 28*(45), 11435-11444.

Friedman, L., Kenny, J. T., Wise, A. L., Wu, D., Stuve, T. A., Miller, D. A., et al. (1998). Brain activation during silent word generation evaluated with functional MRI. *Brain and Language, 64*(2), 231-256.

Frith, C. D. (1992). *The Cognitive Neuropsychology of Schizophrenia*. Hove: Lawrence Erlbaum Associates.

Frith, C. D., & Done, D. J. (1989). Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action. *Psychological Medicine, 19*(2), 359-363.

Garwood, L., Dodgson, G., Bruce, V., & McCarthy-Jones, S. (2013). A preliminary investigation into the existence of a hypervigilance subtype of auditory hallucination in people with psychosis. *Behavioural and Cognitive Psychotherapy, FirstView,* 1-11.

Gawęda, Ł., Woodward, T. S., Moritz, S., & Kokoszka, A. (2013). Impaired action self-monitoring in schizophrenia patients with auditory hallucinations. *Schizophrenia Research, 144*(1–3), 72-79.

George, M. S., Lisanby, S. H., Avery, D., McDonald, W. M., Durkalski, V., Pavlicova, M., et al. (2010). Daily left prefrontal transcranial magnetic stimulation therapy for major depressive disorder: a sham-controlled randomized trial. *Archives of General Psychiatry 67*(5), 507-516.

George, M. S., Wassermann, E. M., Williams, W. A., Callahan, A., Ketter, T. A., Basser, P., et al. (1995). Daily repetitive transcranial magnetic stimulation (rTMS) improves mood in depression. *Neuroreport 6*(14), 1853-1856.

Gould, L. N. (1948). Verbal hallucinations and activity of vocal musculature - an electromyographic study. *American Journal of Psychiatry, 105*(5), 367-372.

Green, M., & Kinsbourne, M. (1990). Subvocal activity and auditory hallucinations - clues for behavioral treatments. *Schizophrenia Bulletin, 16*(4), 617-625.

Green, P., & Preston, M. (1981). Reinforcement of vocal correlates of auditory hallucinations by auditory feedback: a case study. *British Journal of Psychiatry, 139*(3), 204-208.

Hallett, M. (2007). Transcranial magnetic stimulation: a primer. *Neuron, 55*(2), 187-199.

Hoffman, R. E., Boutros, N. N., Berman, R. M., Roessler, E., Belger, A., Krystal, J. H., et al. (1999). Transcranial magnetic stimulation of left temporoparietal cortex in three patients reporting hallucinated "voices". *Biological Psychiatry, 46*(1), 130-132.

Hoffman, R. E., & Cavus, I. (2002). Slow transcranial magnetic stimulation, long-term depotentiation, and brain hyperexcitability disorders. *American Journal of Psychiatry, 159*(7), 1093-1102.

Hoffman, R. E., Fernandez, T., Pittman, B., & Hampson, M. (2011). Elevated functional connectivity along a corticostriatal loop and the mechanism of auditory/verbal hallucinations in patients with schizophrenia. *Biological Psychiatry, 69*(5), 407-414.

Hoffman, R. E., Gueorguieva, R., Hawkins, K. A., Varanko, M., Boutros, N. N., Wu, Y. T., et al. (2005). Temporoparietal transcranial magnetic stimulation for auditory hallucinations: safety, efficacy and moderators in a fifty patient sample. *Biological Psychiatry 58*(2), 97-104.

Hoffman, R. E., Hampson, M., Wu, K., Anderson, A. W., Gore, J. C., Buchanan, R. J., et al. (2007). Probing the pathophysiology of auditory/verbal hallucinations by combining functional magnetic resonance imaging and transcranial magnetic stimulation. *Cerebral Cortex, 17*(11), 2733-2743.

Hoffman, R. E., Rapaport, J., Mazure, C. M., & Quinlan, D. M. (1999). Selective speech perception alterations in schizophrenic patients reporting hallucinated "voices". *American Journal of Psychiatry, 156*(3), 393-399.

Hoffman, R. E., Wu, K., Pittman, B., Cahill, J. D., Hawkins, K. A., Fernandez, T., et al. (2013). Transcranial magnetic stimulation of Wernicke's and right homologous sites to curtail "voices": a randomized trial. *Biological Psychiatry, 73*(10), 1008-1014.

Homan, P., Kindler, J., Hauf, M., Hubl, D., & Dierks, T. (2012). Cerebral blood flow identifies responders to transcranial magnetic stimulation in auditory verbal hallucinations. *Translational Psychiatry, 2*, e189.

Hunter, M. D., Griffiths, T. D., Farrow, T. F. D., Zheng, Y., Wilkinson, I. D., Hegde, N., et al. (2002). A neural basis for the perception of voices in external auditory space. *Brain, 126*(1), 161-169.

Hurlburt, R. T., & Heavey, C. L. (2001). Telling what we know: describing inner experience. *Trends in Cognitive Science, 5*(9), 400-403.

Inouye, T., & Shimizu, A. (1970). Electromyographic study of verbal hallucination. *Journal of Nervous and Mental Disease, 151*(6), 415-422.

Jardri, R., Delevoye-Turrell, Y., Lucas, B., Pins, D., Bulot, V., Delmaire, C., et al. (2009). Clinical practice of rTMS reveals a functional dissociation between agency and hallucinations in schizophrenia. *Neuropsychologia, 47*(1), 132-138.

Jardri, R., Pins, D., Bubrovszky, M., Despretz, P., Pruvo, J. P., Steinling, M., et al. (2007). Self awareness and speech processing: an fMRI study. *NeuroImage, 35*(4), 1645-1653.

Jardri, R., Pouchet, A., Pins, D., & Thomas, P. (2011). Cortical activations during auditory verbal hallucinations in schizophrenia: a coordinate-based meta-analysis. *American Journal of Psychiatry, 168*(1), 73-81.

Johns, L. C., Rossell, S., Frith, C., Ahmad, F., Hemsley, D., Kuipers, E., et al. (2001). Verbal self-monitoring and auditory verbal hallucinations in patients with schizophrenia. *Psychological Medicine, 31*(4), 705-715.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*(1), 3-28.

Jones, C., Hacker, D., Cormac, I., Meaden, A., & Irving., C. B. (2012). Cognitive behaviour therapy versus other psychosocial treatments for schizophrenia. *Cochrane Database Systematic Reviews,* doi: 10.1002/14651858.CD008712.pub2.

Jones, S. R. (2010). Do we need multiple models of auditory verbal hallucinations? Examining the phenomenological fit of cognitive and neurological models. *Schizophrenia Bulletin, 36*(3), 566-575.

Jones, S. R., & Fernyhough, C. (2007). Neural correlates of inner speech and auditory verbal hallucinations: a critical review and theoretical integration. *Clinical Psychology Review, 27*(2), 140-154.

Jones, S. R., & Fernyhough, C. (2007). Thought as action: inner speech, self-monitoring, and auditory verbal hallucinations. *Consciousness and Cognition, 16*(2), 391-399.

Kindler, J., Homan, P., Jann, K., Federspiel, A., Flury, R., Hauf, M., et al. (2013). Reduced neuronal activity in language-related regions after transcranial magnetic stimulation therapy for auditory verbal hallucinations. *Biological Psychiatry 73*(6), 518-524.

Komssi, S., Aronen, H. J., Huttunen, J., Kesäniemi, M., Soinne, L., Nikouline, V. V., et al. (2002). Ipsi- and contralateral EEG reactions to transcranial magnetic stimulation. *Clinical Neurophysiology, 113*(2), 175-184.

Krumbholz, K., Patterson, R. D., Seither-Preisler, A., Lammertmann, C., & Lütkenhöner, B. (2003). Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. *Cerebral Cortex, 13*(7), 765-772.

Langdon, R., Jones, S. R., Connaughton, E., & Fernyhough, C. (2009). The phenomenology of inner speech: comparison of schizophrenia patients with auditory verbal hallucinations and healthy controls. *Psychological Medicine, 39*(4), 655-663.

Larøi, F., Van der Linden, M., & Marczewski, P. (2004). The effects of emotional salience, cognitive effort and meta-cognitive beliefs on a reality monitoring task in hallucination-prone subjects. *British Journal of Clinical Psychology, 43*, 221-233.

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

Lawrie, S. M., Buechel, C., Whalley, H. C., Frith, C. D., Friston, K. J., & Johnstone, E. C. (2002). Reduced frontotemporal functional connectivity in schizophrenia associated with auditory hallucinations. *Biological Psychiatry, 51*(12), 1008-1011.

Lennox, B. R., Park, S. B. G., Medley, I., Morris, P. G., & Jones, P. B. (2000). The functional anatomy of auditory hallucinations in schizophrenia. *Psychiatry Research, 100*(1), 13-20.

Lisanby, S. H., Gutman, D., Luber, B., Schroeder, C., & Sackeim, H. A. (2001). Sham TMS: intracerebral measurement of the induced electrical field and the induction of motor-evoked potentials. *Biological Psychiatry, 49*(5), 460-463.

Looijestijn, J., Diederen, K. M. J., Goekoop, R., Sommer, I. E. C., Daalman, K., Kahn, R. S., et al. (2013). The auditory dorsal stream plays a crucial role in projecting hallucinated voices into external space. *Schizophrenia Research, 146*(1–3), 314-319.

Maeda, F., Keenan, J. P., Tormos, J. M., Topka, H., & Pascual-Leone, A. (2000). Modulation of corticospinal excitability by repetitive transcranial magnetic stimulation. *Clinical Neurophysiology, 111*(5), 800-805.

Mathalon, D. H., Sullivan, E. V., Lim, K. O., & Pfefferbaum, A. (2001). Progressive brain volume changes and the clinical course of schizophrenia in men - a longitudinal magnetic resonance imaging study. *Archives of General Psychiatry, 58*(2), 148-157.

Mathiak, K., Menning, H., Hertrich, I., Mathiak, K. A., Zvyagintsev, M., & Ackermann, H. (2007). Who is telling what from where? A functional magnetic resonance imaging study. *Neuroreport, 18*(5), 405-409.

McCarthy-Jones, S. R., & Fernyhough, C. (2011). The varieties of inner speech: links between quality of inner speech and psychopathological variables in a sample of young adults. *Consciousness and Cognition, 20*(4), 1586-1593.

McGuigan, F. J. (1966). Covert oral behaviour and auditory hallucinations. *Psychophysiology, 3*(1), 73.

McGuire, P. K., David, A. S., Murray, R. M., Frackowiak, R. S. J., Frith, C. D., Wright, I., et al. (1995). Abnormal monitoring of inner speech: a physiological basis for auditory hallucinations. *The Lancet, 346*(8975), 596-600.

McGuire, P. K., Silbersweig, D. A., & Frith, C. D. (1996). Functional neuroanatomy of verbal self-monitoring. *Brain, 119*, 907-917.

McGuire, P. K., Silbersweig, D. A., Murray, R. M., David, A. S., Frackowiak, R. S. J., & Frith, C. D. (1996). Functional anatomy of inner speech and auditory verbal imagery. *Psychological Medicine, 26*(1), 29-38.

Mechelli, A., Allen, P., Amaro, E., Fu, C. H. Y., Williams, S. C. R., Brammer, M. J., et al. (2007). Misattribution of speech and impaired connectivity in patients with auditory verbal hallucinations. *Human Brain Mapping, 28*(11), 1213-1222.

Miniussi, C., Bonato, C., Bignotti, S., Gazzoli, A., Gennarelli, M., Pasqualetti, P., et al. (2005). Repetitive transcranial magnetic stimulation (rTMS) at high and low frequency: an efficacious therapy for major drug-resistant depression? *Clinical Neurophysiology, 116*(5), 1062-1071.

Nayani, T. H., & David, A. S. (1996). The auditory hallucination: a phenomenological survey. *Psychological Medicine, 26*(1), 177-189.

Nitsche, M. A., & Paulus, W. (2000). Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *Journal of Physiology, 527*(3), 633-639.

Nitsche, M. A., & Paulus, W. (2011). Transcranial direct current stimulation - update 2011. *Restorative Neurology and Neuroscience, 29*(6), 463-492.

Opitz, A., Legon, W., Rowlands, A., Bickel, W. K., Paulus, W., & Tyler, W. J. (2013). Physiological observations validate finite element models for estimating subject-specific electric field distributions induced by transcranial magnetic stimulation of the human motor cortex. *NeuroImage, 81*, 253-264.

Pascual-Leone, A., Tormos, J. M., Keenan, J., Tarazona, F., Canete, C., & Catala, M. D. (1998). Study and modulation of human cortical excitability with transcranial magnetic stimulation. *Journal of Clinical Neurophysiology, 15*(4), 333-343.

Petrides, M., & Pandya, D. N. (1988). Association fiber pathways to the frontal cortex from the superior temporal region in the rhesus monkey. *Journal of Comparative Neurology, 273*(1), 52-66.

Price, C. J. (2010). The anatomy of language: a review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences, 1191*(1), 62-88.

Raij, T. T., & Riekki, T. J. J. (2012). Poor supplementary motor area activation differentiates auditory verbal hallucination from imagining the hallucination. *NeuroImage: Clinical, 1*(1), 75-80.

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience, 12*(6), 718-724.

Ridding, M. C., & Rothwell, J. C. (2007). Is there a future for therapeutic use of transcranial magnetic stimulation? *Nature Reviews Neuroscience, 8*(7), 559-567.

Robertson, E. M., Théoret, H., & Pascual-Leone, A. (2003). Studies in cognition: the problems solved and created by transcranial magnetic stimulation. *Journal of Cognitive Neuroscience, 15*(7), 948-960.

Sanfilipo, M., Lafargue, T., Rusinek, H., Arena, L., Loneragan, C., Lautin, A., et al. (2000). Volumetric measure of the frontal and temporal lobe regions in schizophrenia - relationship to negative symptoms. *Archives of General Psychiatry, 57*(5), 471-480.

Sartorius, N., Jablensky, A., Korten, A., Ernberg, G., Anker, M., Cooper, J. E., et al. (1986). Early manifestations and first-contact incidence of schizophrenia in different cultures: a preliminary report on the initial evaluation phase of the WHO collaborative study on determinants of outcome of severe mental disorders. *Psychological Medicine, 16*(04), 909-928.

Schonfeldt-Lecuona, C., Cardenas-Morales, L., Moreno-Aguirre, A., Dorn, K., Langguth, B., Bruhl, A. B., et al. (2012). Effect of 1 Hz repetitive transcranial magnetic stimulation over the auditory cortex on audiometry and otoacoustic emissions. *Brain Topography, 25*(3), 241-247.

Seal, M., Aleman, A., & McGuire, P. (2004). Compelling imagery, unanticipated speech and deceptive memory: neurocognitive models of auditory verbal hallucinations in schizophrenia. *Cognitive Neuropsychiatry 9*(1-2), 43-72.

Shergill, S. S., Bullmore, E. T., Brammer, M. J., Williams, S. C. R., Murray, R. M., & McGuire, P. K. (2001). A functional study of auditory verbal imagery. *Psychological Medicine, 31*(2), 241-253.

Shergill, S. S., Murray, R. M., & McGuire, P. K. (1998). Auditory hallucinations: a review of psychological treatments. *Schizophrenia Research, 32*(3), 137-150.

Siebner, H. R., Lang, N., Rizzo, V., Nitsche, M. A., Paulus, W., Lemon, R. N., et al. (2004). Preconditioning of low-frequency repetitive transcranial magnetic stimulation with transcranial direct current stimulation: evidence for homeostatic plasticity in the human motor cortex. *Journal of Neuroscience, 24*(13), 3379-3385.

Silvanto, J., Muggleton, N. G., Cowey, A., & Walsh, V. (2007). Neural adaptation reveals state-dependent effects of transcranial magnetic stimulation. *European Journal of Neuroscience, 25*(6), 1874-1881.

Silvanto, J., Muggleton, N., & Walsh, V. (2008). State-dependency in brain stimulation studies of perception and cognition. *Trends in Cognitive Sciences, 12*(12), 447-454.

Simons, C. J. P., Tracy, D. K., Sanghera, K. K., O'Daly, O., Gilleen, J., Dominguez, M., et al. (2010). Functional magnetic resonance imaging of inner speech in schizophrenia. *Biological Psychiatry, 67*(3), 232-237.

Silbersweig, D. A., Stern, E., Frith, C., Cahill, C., Holmes, A., Grootoonk, S., et al. (1995). A functional neuroanatomy of hallucinations in schizophrenia. *Nature, 378*, 176-179.

Slotema, C. W., Aleman, A., Daskalakis, Z. J., & Sommer, I. E. (2012). Meta-analysis of repetitive transcranial magnetic stimulation in the treatment of auditory verbal hallucinations: update and effects after one month. *Schizophrenia Research, 142*(1–3), 40-45.

Slotema, C. W., Blom, J. D., de Weijer, A. D., Diederen, K. M., Goekoop, R., Looijestijn, J., et al. (2011). Can low-frequency repetitive transcranial magnetic stimulation really relieve medication-resistant auditory verbal hallucinations? Negative results from a large randomized controlled trial. *Biological Psychiatry, 69*(5), 450-456.

Slotema, C. W., & Daskalakis, J. Z. (2012). Experimental somatic treatments: transcranial magnetic stimulation in the treatment of auditory verbal hallucinations - a meta-analysis and review. In J. D. Blom & I. Sommer (Eds.), *Hallucinations: Research and Practice*. New York: Springer.

Sommer, I. E. C., Diederen, K. M. J., Blom, J. D., Willems, A., Kushan, L., Slotema, K., et al. (2008). Auditory verbal hallucinations predominantly activate the right inferior frontal area. *Brain, 131*, 3169-3177.

Sommer, I. E. C., Slotema, C. W., Daskalakis, Z. J., Derks, E. M., Blom, J. D., & van der Gaag, M. (2012). The treatment of hallucinations in schizophrenia spectrum disorders. *Schizophrenia Bulletin, 38*(4), 704-714.

Stagg, C. J., & Nitsche, M. A. (2011). Physiological basis of transcranial direct current stimulation. *Neuroscientist, 17*(1), 37-53.

Tien, A. Y. (1991). Distribution of hallucinations in the population. *Social Psychiatry & Psychiatric Epidemiology, 26*(6), 287-292.

Vercammen, A., Knegtering, H., Bruggeman, R., & Aleman, A. (2010). Subjective loudness and reality of auditory verbal hallucinations and activation of the inner speech processing network. *Schizophrenia Bulletin, 37*(5), 1009-1016.

Vercammen, A., Knegtering, H., den Boer, J. A., Liemburg, E. J., & Aleman, A. (2010). Auditory hallucinations in schizophrenia are associated with reduced functional connectivity of the temporo-parietal area. *Biological Psychiatry, 67*(10), 912-918.

Vercammen, A., Knegtering, H., Liemburg, E. J., den Boer, J. A., & Aleman, A. (2010). Functional connectivity of the temporo-parietal region in schizophrenia: effects of rTMS treatment of auditory hallucinations. *Journal of Psychiatric Research, 44*(11), 725-731.

Walsh, V., & Cowey, A. (2000). Transcranial magnetic stimulation and cognitive neuroscience. *Nature Reviews Neuroscience, 1*(1), 73-79.

Walsh, V., & Pascual-Leone, A. (2003). *Transcranial Magnetic Stimulation: A Neurochronometrics of Mind*. Cambridge, Massachusetts: The MIT Press.

Wassermann, E. M., Wedegaertner, F. R., Ziemann, U., George, M. S., & Chen, R. (1998). Crossed reduction of human motor cortex excitability by 1-Hz transcranial magnetic stimulation. *Neuroscience Letters, 250*(3), 141-144.

Waters, F., Allen, P., Aleman, A., Fernyhough, C., Woodward, T. S., Badcock, J. C., et al. (2012). Auditory hallucinations in schizophrenia and nonschizophrenia populations: a review and integrated model of cognitive mechanisms. *Schizophrenia Bulletin, 38*(4), 683-692.

Waters, F., Badcock, J., Michie, P., & Maybery, M. (2006). Auditory hallucinations in schizophrenia: intrusive thoughts and forgotten memories. *Cognitive Neuropsychiatry, 11*(1), 65-83.

Waters, F., Woodward, T., Allen, P., Aleman, A., & Sommer, I. (2012). Self-recognition deficits in schizophrenia patients with auditory hallucinations: a meta-analysis of the literature. *Schizophrenia Bulletin, 38*(4), 741-750.

Whitford, T. J., Ford, J. M., Mathalon, D. H., Kubicki, M., & Shenton, M. E. (2012). Schizophrenia, myelination, and delayed corollary discharges: a hypothesis. *Schizophrenia Bulletin, 38*(3), 486-494.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science, 269*(5232), 1880-1882.

Wykes, T., Steel, C., Everitt, B., & Tarrier, N. (2008). Cognitive behavior therapy for schizophrenia: effect sizes, clinical models, and methodological rigor. *Schizophrenia Bulletin, 34*(3), 523-537.

Young, L., Camprodon, J. A., Hauser, M., Pascual-Leone, A., & Saxe, R. (2010). Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proceedings of the National Academy of Science USA, 107*(15), 6753-6758.

Young, L., Dodell-Feder, D., & Saxe, R. (2010). What gets the attention of the temporo-parietal junction? An fMRI investigation of attention and theory of mind. *Neuropsychologia, 48*(9), 2658-2664.

Chapter 2

Methodological review and experimental aims of this thesis

Chapter 1 provided a review of contemporary cognitive neuroscientific theories of auditory verbal hallucinations (AVHs), with a particular focus on attempts to treat AVHs with neurostimulation techniques. The following chapter will give an overview of the theoretical and technical methodology used as part of this thesis, and summarise the four experimental chapters that follow. Firstly, a brief overview of signal detection theory and its applicability to the study of hallucinations will be given. Much of the information in Section 1.1 is from Stanislaw and Todorov (1999), although the examples of signal detection theory usage for task performance analysis are based on tasks more relevant to the study of hallucinations. This is followed by a summary of the relative advantages and disadvantages of noninvasive neurostimulation techniques which are used in two of the experimental chapters, with a justification for usage of the technique in this thesis.

1. Signal detection theory

1.1. Overview

Signal detection theory (SDT) can be used whenever a task requires participants to distinguish between two types of stimulus – for example, whether they believe a stimulus to be present or absent, or whether they believe a perception was self-generated or non-self-generated. For instance, if an individual must decide whether they perceived a voice stimulus in random noise, there are four possible outcomes, depending on the presence or absence of a stimulus, and their subsequent response. In the presence of the stimulus, the individual may 1) correctly respond that they believe a voice to be present (a 'hit', or a true positive), or 2) incorrectly respond that they do not believe a voice is present (a 'miss', or a false negative). Alternatively, in the absence of a voice, the individual may 3) correctly respond that a voice is not present (a 'correct rejection', or a true negative), or 4) incorrectly respond that they believe a voice to be present (a 'false alarm', or a false positive). The key dependent variables in a signal detection task are therefore the 'hit rate' (the number of hits, divided by the total number of 'stimulus present' trials), and the 'false alarm rate' (the number of false alarms, divided by the number of 'stimulus absent' trials). These can also be expressed as a percentage, if multiplied by 100. (The 'miss rate' and 'correct rejection rate' are generally not discussed, since they would be directly proportional to the hit rate and false alarm rate, respectively). Furthermore, hit rates and false alarm rates can be used to quantify the capability to distinguish between signal and noise (sensitivity), and the criterion which individuals use to accept the presence of a signal (response bias).

Sensitivity, denoted by d', is defined as the difference between the standardised hit rate and the standardised false alarm rate[1]. Thus, if a participant correctly detected the signal on the majority of signal-present trials (a high hit rate), and rarely made false alarm responses (a low false alarm rate), there would be a large difference between the hit and false alarm rate, and therefore the participant would have a relatively high d' value (that is, they have high task sensitivity). Conversely, if the participant struggled to detect the signal when it was present, and also frequently incorrectly detected the signal (therefore having a high false alarm rate), they would have a relatively low d' value, because there would be little difference between the hit and false alarm rate. d' is therefore a useful measure of ability to distinguish signal from noise.

Response bias (denoted by β, sometimes also referred to as 'response criterion') refers to the tendency of the participant to respond 'yes' or 'no' on the task. For example, a participant with a low response bias would be more willing to respond 'yes' on a signal detection task, regardless of the presence or absence of a signal, and would therefore be likely to have a high hit rate, and a high false alarm rate. Conversely, an individual with a high response bias would be less willing to respond 'yes', and would therefore have a low hit rate, and a low false alarm rate. Throughout this thesis, response bias will be calculated as outlined in Stanislaw and Todorov (1999), as follows: $\beta = e\left\{\frac{Z(FA)^2 - Z(H)^2}{2}\right\}$, where *FA* represents false alarm rate, *H* represents hit rate, *z* represents standardised scores (i.e., *z*-scores), and *e* represents exponentiation.

---

[1] Nonparametric measures of sensitivity such as *A'* also exist, although are not used in this thesis, and so are not discussed any further. Likewise, nonparametric versions of response bias also exist, but will not be discussed any further.

1.2. Signal detection theory and hallucinations

Clearly, of interest in the study of hallucinations is the propensity to make false alarm responses, i.e., to report that a signal is present, when one is not. A question of interest is therefore whether hallucinations (conceptualised as false alarms) are due to low perceptual sensitivity, or a low response bias. Interestingly, these two possibilities would make opposite predictions about performance in the presence of a signal: that is, if hallucinations were due to low perceptual sensitivity, they would likely be accompanied by deficits in detection of presented perceptual stimuli; however, if hallucinations were due to a low response bias, they would be accompanied by a higher detection of presented perceptual stimuli.

Previously, this issue has been investigated in samples with diagnoses of schizophrenia, comparing those that hallucinate and those that do not (Brookwell, Bentall, & Varese, 2013). In a typical auditory signal detection task, participants are asked to listen to bursts of noise through headphones, and, for each burst, asked to respond whether they believed a voice was present. Importantly, on some trials a voice is presented, usually at a pre-determined level likely to be close to the participant's auditory threshold, and on other trials, no voice is presented. (The volume and signal-to-noise ratio of the stimuli must be kept constant across all participants, however, for signal detection parameters to be comparable across the sample.) In clinical samples, individuals who hallucinate tend to have a reduced response bias (β), but similar levels of sensitivity (d'), compared to individuals who do not hallucinate, or healthy controls (Bentall & Slade, 1985; Varese, Barkus, & Bentall, 2012; though see Vercammen, de Haan, & Aleman, 2008, for an exception). This finding has also been extended to non-clinical samples, showing that individuals who report higher levels of hallucination-proneness showed a

63

reduced response bias, but no change in sensitivity (Barkus et al., 2011; Varese, Barkus, & Bentall, 2011; see Brookwell, Bentall, & Varese, 2013, for a meta-analysis). Thus, individuals who hallucinate, or who seem to be more prone to such experiences, show a style of responding in which they are more willing to label an ambiguous percept as 'signal present'. An alternative interpretation of these findings would be that highly hallucination-prone individuals are more willing to accept that an internally generated percept is external (i.e., has appeared in the noise).

This alternative interpretation is supported by a parallel stream of research which has investigated memory for the source of information. A common paradigm, based on the source monitoring framework, presents participants with a series of words, half of which they are required to read aloud themselves, and half of which the experimenter reads aloud. At a later point, the participant is then required to recall which words were self-generated, and which were non-self-generated (Johnson, Hashtroudi, & Lindsay, 1993). Here, signal detection theory is again useful in the data analysis; instances in which the participant correctly recalls a word as read aloud by the experimenter can be conceptualised as a 'hit' (because they have correctly identified an 'external' signal as present), and instances in which the participant incorrectly recalls a word as read aloud by the experimenter can be conceptualised as a 'false alarm' (because they have incorrectly labelled an item that was self-generated as externally generated). Similarly to the auditory paradigm discussed above, a common finding in both clinical and non-clinical samples is that individuals who are more prone to hallucinations show a lower response bias (they are more willing to respond that a word was non-self-generated, regardless of its source), but comparable levels of sensitivity (that is, they do not seem to show a general deficit in source memory performance) (Bentall, Baker, & Havers, 1991;

Brookwell, et al., 2013; Morrison & Haddock, 1997; Woodward, Menon, & Whitman, 2007). This evidence therefore suggests that the lowered response bias seen in individuals prone to hallucinations may be linked to a propensity to label perceptions as external/non-self-generated, both in the auditory signal detection paradigm and the source memory paradigm.

As discussed in Chapter 1, the most prominent cognitive neuroscientific model of self-monitoring suggests that a forward model is used to predict the sensory consequences of self-produced action, attenuating activity in sensory cortices in response to the resulting incoming perceptual information. According to this model, when a motor command is sent, an efference copy of the command is sent to sensory cortices, which is compared with incoming sensory information; if there is a mismatch, the sensory information may be experienced as non-self-generated (Wolpert, Ghahramani, & Jordan, 1995). Therefore, it has been suggested that a disrupted or delayed efference copy of a speech motor command could lead to inner speech being misattributed to an external source (Frith, Blakemore, & Wolpert, 2000; Jones & Fernyhough, 2007; Seal, Aleman, & McGuire, 2004). A disrupted efference copy signal would increase the likelihood that self-generated actions would erroneously be experienced as non-self-generated (but would not cause non-self-generated actions to be experienced as self-generated, since perception of non-self-produced actions should not be affected by a disrupted efference copy mechanism). Therefore, in signal detection terminology, a disrupted efference copy might lead to a reduced response bias in decisions regarding whether a perception was self-generated or not.

A separate, though related, strand of research has attempted to explain how internal mental events may be experienced as external, through the source monitoring

framework (SMF). Described in more detail in Chapters 4 and 6, the SMF would also make predictions regarding signal detection response biases in tasks that require participants to monitor their own actions. The SMF specifies that self- or non-self-generated events are experienced as such by reference to subjective qualities such as recalled vividness, cognitive effort associated with the recollection, and affective or semantic content (Johnson et al., 1993). As such, aspects of inner experience such as heightened perceptual content/vividness, or low cognitive effort, may lead to a higher likelihood of external misattributions (but are unlikely to lead to incorrectly misattributing an external event as internal). The SMF would therefore also predict that hallucinations may be associated with a response bias towards responding that events took place externally, rather than a change in task sensitivity.

It should be noted that efference copy and SMF theories are not mutually exclusive; for example, it is possible that atypical efference copy mechanisms may be linked to the cognitive effort or vividness of mental imagery, although this possibility has not yet been empirically tested. The issue of the overlap between different cognitive explanations of self-/source-monitoring is addressed in Chapter 4.

2. Transcranial direct current stimulation

Chapters 5 and 6 of this thesis investigate the role of specific cortical regions in some of the tasks discussed throughout Chapters 1 and 2. The technique employed to investigate this, transcranial direct current stimulation (tDCS), is a neurostimulation method which can transiently increase or decrease the level of cortical excitability in neurons underneath electrodes placed on the scalp. As such, it is an ideal technique

to investigate the effects of attenuating or increasing cortical activity on signal detection biases, as described above.

Using tDCS involves running a weak electrical current (usually < 2.0mA, with a current density of < 0.06 mA/cm^2) between two electrodes which are placed on the participant's scalp. Typically, stimulation increases excitability underneath the anodal electrode, and decreases excitability underneath the cathodal electrode; for example, anodal stimulation of motor cortex reduces the motor threshold (i.e., reduces the strength of cortical stimulation required to evoke a hand movement), whilst cathodal stimulation has the opposite effect (Nitsche et al., 2008; Nitsche & Paulus, 2000; Stagg & Nitsche, 2011). During stimulation, cortical effects seem to be due to hyperpolarisation (cathodal) or depolarisation (anodal) of neurons underlying the electrode (Nitsche & Paulus, 2000). After-effects of stimulation have also been shown, both at a neural (Ardolino, Bossi, Barbieri, & Priori, 2005) and behavioural level (Hummel & Cohen, 2006). This appears to be due to a reduction in GABA concentration underneath the anodal electrode and a reduction in glutamate concentration underneath the cathodal electrode (Stagg & Nitsche, 2011).

The main advantage of using tDCS is that it allows causal inferences regarding the role of specific brain areas in cognition. Using tDCS, it is possible to modulate activity in cortical regions and then observe the resultant changes in task performance, which in turn is evidence that activity in that brain area plays a causal role in performing the task. This is in contrast to techniques such as functional magnetic resonance imaging (fMRI), which can show correlations between levels of activity in certain brain areas and performance on specific tasks, but only allows a limited level of insight into the direction of causation (observed activity in a certain brain area could, for example, be epiphenomenal to the task being performed). As

such, a common research trajectory is to use neuroimaging techniques such as fMRI

to investigate BOLD activations during task performance (which may implicate

multiple brain areas), and then to test the causal role of these areas using

neurostimulation techniques. For example, neuroimaging evidence suggested that

posterior parietal cortex plays an important role in visual search tasks for conjunction

items (Shafritz, Gore, & Marois, 2002). This was then confirmed using TMS to

interrupt functioning in the right posterior parietal cortex, and observing an increase

in reaction time for a conjunction visual search task (Ellison, Lane, & Schenk, 2007).

Indeed, a similar approach has been taken in this thesis, in that neuroimaging

evidence suggests that superior temporal cortical regions (among a number of other

regions) may be important in biases in auditory signal detection and source memory

(Barkus, Stirling, Hopkins, McKie, & Lewis, 2007; Mitchell & Johnson, 2009;

Sugimori, Mitchell, Raye, Greene, & Johnson, 2014). The studies reported here

apply tDCS to superior temporal regions on the basis of previous findings suggesting

that high levels of activation in these regions may cause misattribution of self-

generated perceptions to a non-self source.

Logistically, tDCS is also an ideal technique to apply during tasks which require

participants to attend to auditory stimuli, because it does not cause any noise itself

(as opposed to other neurostimulation techniques such as transcranial magnetic

stimulation (TMS), which emit loud clicking noises). Furthermore, with tDCS,

'sham' stimulation is an effective and simple control condition against which to

compare any effects of stimulation (Gandiga, Hummel, & Cohen, 2006). Typically,

sham stimulation involves stimulating for a much shorter time period (usually ~30

secs), without telling the participant that the stimulation will stop earlier than they

were originally informed. This length of stimulation is not sufficient to cause

significant neuronal modulation, but does elicit the tingling sensation characteristic of tDCS underneath the electrodes. Sham tDCS is therefore thought to be an effective method of blinding participants to the experimental condition (Gandiga, et al., 2006) (see Chapter 1, Section 4.1 for a discussion of this issue).

One drawback to using tDCS, however, is its relative lack of spatial resolution: the sizes of the stimulating electrodes vary between studies, but are typically between 25–35cm^2. Coupled with the possibility of spread of activation between brain regions following neurostimulation (Pascual-Leone, Tormos, Keenan, Tarazona, Canete, & Catala, 1998), this means that firm conclusions about the precise areas being stimulated are hard to draw without concurrent neuroimaging techniques (see Ellison et al., 2014, for an example). Nevertheless, given that neurostimulation research typically draws on neuroimaging studies that have identified activity in relatively precise cortical regions, it is often parsimonious to conclude that any behavioural effects are due to stimulating the previously identified cortical regions.

3. Experimental aims, and outline of empirical studies in this thesis

Broadly, the studies reported in this thesis aim to explore the cognitive and neural substrates of hallucinatory experiences in a non-clinical sample. As outlined above, the key techniques used to accomplish this are tasks amenable to analysis using signal detection theory, and noninvasive brain stimulation techniques such as tDCS, used to investigate the role of superior temporal regions in these tasks.

Chapter 3 aims to explore the role that auditory verbal mental imagery plays in signal detection, testing the prediction that engaging participants in mental imagery

will result in a lower response bias, particularly in participants prone to hallucinatory experiences. Chapter 4 extends the findings from that study, using hierarchical regression analyses to investigate whether a variety of different tasks which putatively require the participants to monitor their own actions predict self-reported hallucination-proneness. Reported phenomenological differences in the use of inner speech are also studied in relation to hallucination-proneness. Drawing on the research discussed and the conclusions drawn in Chapter 1, Chapter 5 then uses tDCS to test the role of the superior temporal lobe in auditory signal detection, whilst Chapter 6 uses a similar methodology to test the role of superior temporal and prefrontal cortices in source memory.

References

Ardolino, G., Bossi, B., Barbieri, S., & Priori, A. (2005). Non-synaptic mechanisms underlie the after-effects of cathodal transcutaneous direct current stimulation of the human brain. *Journal of Physiology, 568*(2), 653-663.

Barkus, E., Smallman, R., Royle, N., Barkus, C., Lewis, S., & Rushe, T. (2011). Auditory false perceptions are mediated by psychosis risk factors. *Cognitive Neuropsychiatry, 16*(4), 289-302.

Barkus, E., Stirling, J., Hopkins, R., McKie, S., & Lewis, S. (2007). Cognitive and neural processes in non-clinical auditory hallucinations. *British Journal of Psychiatry, 191*, S76-S81.

Bentall, R. P., Baker, G. A., & Havers, S. (1991). Reality monitoring and psychotic hallucinations. *British Journal of Clinical Psychology, 30*, 213-222.

Bentall, R. P., & Slade, P. D. (1985). Reality testing and auditory hallucinations - a signal detection analysis. *British Journal of Clinical Psychology, 24*, 159-169.

Brookwell, M. L., Bentall, R. P., & Varese, F. (2013). Externalizing biases and hallucinations in source-monitoring, self-monitoring and signal detection studies: a meta-analytic review. *Psychological Medicine, 43*(12), 2465-2475.

Ellison, A., Ball, K. L., Moseley, P., Dowsett, J., Smith, D. T., Weis, S., et al. (2014). Functional interaction between right parietal and bilateral frontal cortices during visual search tasks revealed using functional magnetic imaging and transcranial direct current stimulation. *PLoS ONE, 9*(4), e93767.

Ellison, A., Lane, A. R., & Schenk, T. (2007). The interaction of brain regions during visual search processing as revealed by transcranial magnetic stimulation. *Cerebral Cortex, 17*(11), 2579-2584.

Frith, C. D., Blakemore, S.-J., & Wolpert, D. M. (2000). Explaining the symptoms of schizophrenia: abnormalities in the awareness of action. *Brain Research Reviews, 31*(2–3), 357-363.

Gandiga, P. C., Hummel, F. C., & Cohen, L. G. (2006). Transcranial DC stimulation (tDCS): a tool for double-blind sham-controlled clinical studies in brain stimulation. *Clinical Neurophysiology, 117*(4), 845-850.

Hummel, F. C., & Cohen, L. G. (2006). Non-invasive brain stimulation: a new strategy to improve neurorehabilitation after stroke? *Lancet Neurology, 5*(8), 708-712.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*(1), 3-28.

Jones, S. R., & Fernyhough, C. (2007). Thought as action: inner speech, self-monitoring, and auditory verbal hallucinations. *Consciousness and Cognition, 16*(2), 391-399.

Mitchell, K. J., & Johnson, M. K. (2009). Source monitoring 15 years later: what have we learned from fMRI about the neural mechanisms of source memory? *Psychological Bulletin, 135*(4), 638-677.

Morrison, A. P., & Haddock, G. (1997). Cognitive factors in source monitoring and auditory hallucinations. *Psychological Medicine, 27*(3), 669-679.

Nitsche, M. A., Cohen, L. G., Wassermann, E. M., Priori, A., Lang, N., Antal, A., et al. (2008). Transcranial direct current stimulation: state of the art 2008. *Brain Stimulation, 1*(3), 206-223.

Nitsche, M. A., & Paulus, W. (2000). Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *The Journal of Physiology, 527*(3), 633-639.

Pascual-Leone, A., Tormos, J. M., Keenan, J., Tarazona, F., Canete, C., & Catala, M. D. (1998). Study and modulation of human cortical excitability with transcranial magnetic stimulation. *Journal of Clinical Neurophysiology, 15*(4), 333-343.

Seal, M., Aleman, A., & McGuire, P. (2004). Compelling imagery, unanticipated speech and deceptive memory: neurocognitive models of auditory verbal hallucinations in schizophrenia. *Cognitive Neuropsychiatry, 9*(1-2), 43-72.

Shafritz, K. M., Gore, J. C., & Marois, R. (2002). The role of the parietal cortex in visual feature binding. *Proceedings of the National Academy of Sciences USA, 99*(16), 10917-10922.

Stagg, C. J., & Nitsche, M. A. (2011). Physiological basis of transcranial direct current stimulation. *Neuroscientist, 17*(1), 37-53.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137-149.

Sugimori, E., Mitchell, K. J., Raye, C. L., Greene, E. J., & Johnson, M. K. (2014). Brain mechanisms underlying reality monitoring for heard and imagined words. *Psychological Science, 25*(2), 403-413.

Varese, F., Barkus, E., & Bentall, R. P. (2011). Dissociative and metacognitive factors in hallucination-proneness when controlling for comorbid symptoms. *Cognitive Neuropsychiatry, 16*(3), 193-217.

Varese, F., Barkus, E., & Bentall, R. P. (2012). Dissociation mediates the relationship between childhood trauma and hallucination-proneness. *Psychological Medicine, 42*(5), 1025-1036.

Vercammen, A., de Haan, E. H. F., & Aleman, A. (2008). Hearing a voice in the noise: auditory hallucinations and speech perception. *Psychological Medicine, 38*(8), 1177-1184.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science, 269*(5232), 1880-1882.

Woodward, T. S., Menon, M., & Whitman, J. C. (2007). Source monitoring biases and auditory hallucinations. *Cognitive Neuropsychiatry, 12*(6), 477-494.

Chapter 3

Biased monitoring of inner speech in hallucination-prone individuals: the effect of auditory verbal imagery on signal detection

Abstract

Cognitive models of auditory verbal hallucinations have suggested that they may result when internal mental events, such as inner speech or auditory verbal imagery (AVI), are misattributed to an external source. This has been supported by numerous studies indicating that individuals who experience hallucinations tend to perform in a biased manner on tasks that require them to distinguish self-generated from non-self-generated perceptions. However, these tasks have typically been of limited relevance to inner speech or hallucinations. Here, a paradigm is used in which participants were instructed to use AVI whilst completing an auditory signal detection task. It was hypothesized that AVI-usage would cause participants to perform in a biased manner, therefore falsely detecting more voices in bursts of noise. In a first experiment, when cued to generate AVI, highly hallucination-prone participants showed a lower response bias than when performing a standard signal detection task, whereas participants not prone to hallucinations performed no differently between the two conditions. In a second experiment, participants were not specifically instructed to use AVI, but highly hallucination-prone participants who retrospectively reported using AVI showed the same lowered response bias. Results are discussed in relation to prominent inner speech models of hallucinations.

1. Introduction

1.1. Auditory verbal hallucinations and inner speech

Auditory verbal hallucinations (AVHs) are the experience of hearing a voice in the absence of any speaker. Although commonly associated with a diagnosis of schizophrenia, AVHs also occur in around 1.5–3% of the healthy, nonclinical population (Tien, 1991). There is emerging evidence that the predisposition to AVHs may lie on a continuum, ranging from individuals who frequently experience, to individuals who rarely or never report, hallucinations (Johns et al., 2014; Johns & van Os, 2001). A fruitful area of investigation is therefore to investigate whether cognitive traits and biases associated with hallucinations in clinical populations are shared by individuals in the general population who report frequent hallucinatory experiences.

The most prominent cognitive model of AVHs suggests that they occur when an internal mental event (such as inner speech or auditory verbal imagery – AVI) is misattributed to an external source (Ditman & Kuperberg, 2005; Frith, 1992; Jones & Fernyhough, 2007b). This strand of research has therefore been embedded in the source monitoring framework, which attempts to explain how we make judgements regarding the origin of information (i.e., its source) (Johnson, Hashtroudi, & Lindsay, 1993). Specifically, a bias in 'reality monitoring', which refers to the ability to distinguish between internally generated and externally generated perceptions, has been linked to AVHs (Bentall, Baker, & Havers, 1991). Externalising biases have variously been linked to excessively vivid mental imagery (Aleman, Böcker, Hijman, de Haan, & Kahn, 2003), and low cognitive effort/intrusiveness associated with mental imagery (Jones & Fernyhough, 2009; Morrison, Haddock, & Tarrier,

1995). On a mechanistic level, forward models may be involved in predicting the sensory consequences of motor processes, and successful prediction via an efference copy may be one way in which self-generated actions are experienced as such (Frith, Blakemore, & Wolpert, 2000). Aberrant efference copy mechanisms could therefore underlie the external misattribution of internal mental events (Ford & Mathalon, 2005).

Reality monitoring for verbal stimuli has typically been assessed using source memory paradigms, which require participants to recall whether words were spoken by the experimenter or by themselves. A common finding is that patients with a diagnosis of schizophrenia who hallucinate, compared to those who do not hallucinate or healthy control participants, are more likely to misremember words as having been spoken by the experimenter (Woodward, Menon, & Whitman, 2007), therefore showing an 'externalising bias' on reality monitoring tasks. Providing support for continuum models of AVHs, non-clinical samples who report higher levels of hallucination-proneness also show a similar pattern of responding on reality monitoring tasks (Brookwell, Bentall, & Varese, 2013; Larøi, van der Linden, & Marczewski, 2004).

A similar line of research has attempted to engage the participant in an 'online' decision making process, referred to as 'reality discrimination', requiring participants to immediately respond as to whether a perception was internal or external (in contrast to the 'offline' decisions required in a source memory task, which typically require a decision to be made at a later time point, e.g., Woodward, et al., 2007). Reality discrimination tasks typically take the form of signal detection tasks, in which the participant must decide whether a voice was present in a burst of noise. In these tasks, hallucinating patients tend to show a bias towards responding

that a voice is present in the noise (Varese, Barkus, & Bentall, 2012). In a non-clinical sample, participants who reported more hallucinatory experiences also showed the same bias in responding (Barkus et al., 2011) (see Chapter 2 for a discussion of signal detection methodology). These findings have been linked theoretically to the reality monitoring tasks described above, as providing evidence linking AVHs to an externalising bias (Brookwell, Bentall, & Varese, 2013).

However, these tasks are not ideally positioned to test models of AVHs that specify the misattribution of internal mental events such as inner speech, for two main reasons: 1) they are not 'online' measures (source memory tasks, for example, are 'offline' in that they require participants to decide who generated words earlier in the testing session); 2) they are either not specific to monitoring of speech or, if they are, are likely to use 'overt' (out loud) speech, as opposed to engaging the participant in auditory verbal imagery or inner speech. This limits the applicability of the results to inner speech models of AVHs, because it assumes that overt vocalisation in an experimental situation utilises the same mechanisms as covert or inner speech. Although there is evidence that overt and covert speech share cognitive and neural mechanisms, particularly in relation to the motor system (Perrone-Bertolotti, Rapin, Lachaux, Baciu, & Lœvenbruck, 2014), applicability of findings from studies requiring participants to use overt speech is limited.

A more valid paradigm would require participants to engage in covert auditory verbal mental imagery, whilst simultaneously detecting the presence or absence of a similar auditory verbal stimulus, thus sidestepping both of the above problems. This would engage participants in covert auditory imagery, giving them the opportunity to misattribute internally generated imagery to an external source (i.e., falsely respond

that they believed a stimulus was presented), without including a time interval between stimulus presentation and participant response.

1.2. Mental imagery and perception

One task-based measure that meets the criteria outlined above (requiring an 'online' response from the participant, and using auditory imagery) has been used to study the interaction between mental imagery and perception. Perky (1910), a student in Edward Titchener's laboratory which mainly used methods involving introspection by the participant, was the first to systematically study the effect of mental imagery on perception. Perky carried out a series of experiments that suggested that visual imagery interfered with the simultaneous perception of a visually presented stimulus (subsequently referred to as the Perky Effect). This was taken to indicate that, since mental imagery and perception could be confused, they must rely on similar mechanisms. However, others have found that mental imagery actually facilitates perceptions in the visual modality (Peterson & Graham, 1974). This finding has also been replicated in the auditory modality; for example, Farah and Smith (1983) engaged participants in auditory imagery of a pure tone, whilst simultaneously requiring them to detect a similar tone in noise. Participants were therefore required to distinguish between self-generated, internal mental imagery and an external stimulus. The results showed that using auditory imagery facilitated perception of the tone, although the task used did not include trials with no signal present, and so signal detection analysis was not reported. Findings on the interaction between imagery and perception have, therefore, been equivocal.

More recently, Aleman et al. (2003) used a similar paradigm with a sample of patients with a diagnosis of schizophrenia, showing that the 'gain' on perception of a

pure tone due to auditory imagery was strongly correlated with hallucination severity. This finding was interpreted as reflecting an over-reliance on top-down processes in hallucinating patients (which could also be related to a bias towards labelling internal imagery as external). It should be noted that there was not a significant difference in performance on this task between non-hallucinating and hallucinating patients. This may have been due to the criterion used to separate hallucinating from non-hallucinating patients in the analysis, which meant that some patients in the non-hallucinating group may have hallucinated as recently as three months before participating.

One problem with these studies is that they do not measure the effect of imagery on the tendency to *falsely* detect a signal in noise[2], because there is always a signal present. This is a key variable when linking performance to the tendency to hallucinate, and also when performing signal detection analysis. From the data presented by Aleman et al. (2003), for example, it is not possible to tell whether the 'gain' on perception was due to a change in sensitivity (an increased ability to distinguish signal from noise), or a change in response bias (i.e., participants being more willing to respond that a tone was present when using imagery). The previously discussed literature relating to biases in reality monitoring/discrimination would imply that it may be the latter. Imagery-perception interaction tasks, though, have the advantage of directly engaging participants in internal mental imagery (as opposed to speaking aloud), and requiring them to distinguish whether any subsequent perception was internally generated or not, hence meeting the criteria

---

[2] Farah & Smith (1983), however, note that many participants, when using auditory imagery of a pure tone, reported hearing the tone at times discordant with when the tone was actually presented, implying that imagery may have cause false detections. These false alarm responses were not quantified, however.

outlined above. As well as being informative on the nature of mental imagery, this makes the tasks ideal for testing inner speech models of AVHs.

A further question stemming from inner speech models of AVHs relates to what may cause one instance of inner speech to become misattributed, but not another instance. As discussed, source monitoring theories may appeal to vividness of mental imagery and the effort associated with cognitions, but theories of AVHs have also suggested that unpleasant, ego-dystonic cognitions are more likely to become misattributed (Morrison, et al., 1995). This is consistent with reports that AVHs are more likely to occur in conditions of negative affect or stress (Nayani & David, 1996). However, the findings of studies that have investigated whether negatively valenced cognitions are more likely to become externally misattributed than neutral or positively-valenced cognitions have been inconsistent (Bendall, Jackson, & Hulbert, 2011; Morrison, et al., 1995).

In a recent meta-analysis of these studies (Brookwell et al., 2013), slightly larger externalizing biases were observed in studies that employed positive (Hedges g = 0.75), than negative (Hedges g = 0.62), or than neutral stimuli (Hedges g = 0.50). However, because of the designs of the synthesized studies, it was not possible to examine whether these effect sizes differed from one another, which makes drawing any conclusions difficult (especially when only a small number of studies (five) were included in the meta-analysis). Recent work investigating how emotion can modulate intentional binding (a low-level measure of sensorimotor agency) showed that intentional binding was reduced (i.e., participants' sense of agency over their actions was reduced) when participants' actions were paired with a *negative* emotional outcome in comparison to when actions were paired with a neutral or positive outcome (Yoshie & Haggard, 2013). Thus, there are reasons to believe that

79

negatively-valenced cognitions may be more likely to become misattributed to an external source than are neutral or positively-valenced cognitions, although previous research has been equivocal.

The present study therefore set out to test whether performance on an auditory signal detection task was affected by the generation of auditory verbal mental imagery, and whether this was linked to the emotional valence of the stimuli, and/or self-reported hallucination-proneness, in a non-clinical sample. However, two key changes were made to previous paradigms investigating AVI and signal detection. Firstly, trials in which no signal was present were included, to allow the generation of 'false alarm' responses and the calculation of relevant signal detection measures. Secondly, the signal detected was a voice (as opposed to a pure tone), in order to maximise the relevance of the task to inner speech models of AVHs.

We conducted two experiments which manipulated or measured the extent to which participants generated AVI. In the first experiment, participants completed two blocks of auditory signal detection: one in which they were cued with a short sentence and required to generate AVI of that sentence whilst performing the task, and one in which there was no cued sentence or instruction to use AVI. In the second experiment, participants completed the same two blocks, but with no instruction to use AVI; instead, participants were invited retrospectively to report on the extent to which they felt they had used AVI whilst attempting to detect the auditory stimulus. We predicted that use of auditory verbal imagery would lower participants' response bias, because it would lead to a higher likelihood of external misattributions (i.e., participants would have more opportunity to misattribute a self-generated event to an external source). Such a bias would lead participants to correctly detect more voices in the noise, but also incorrectly detect more voices in the noise (i.e., report hearing a

voice when none was present). Furthermore, drawing on the previous literature linking imagery-perception interactions and AVHs, we predicted that participants who scored highly on self-report measures of proneness to hallucinations may be particularly vulnerable to this effect. Finally, consistent with the findings of Yoshie and Haggard (2013) but in contrast to the findings of Brookwell et al. (2013), we predicted that participants' response biases would be lower when they generated negative AVI than when they generated positive AVI.

2. Experiment 1

In the first experiment, participants performed an auditory signal detection task under two conditions: 1) with a visually presented verbal cue using AVI and 2) with no cue, and no instruction to use AVI. We predicted that using imagery would lower participants' response bias (making them more likely to report hearing a voice in the noise), but have no effect on sensitivity (the ability to distinguish between the voice and the noise).

2.1. Materials and Methods

2.1.1. Participants

The sample consisted of 125 participants from the undergraduate and staff population of Durham University, UK. Five participants were excluded from the final sample, due to technical malfunctions during the testing session ($n = 3$), or because their task sensitivity (d' – see below) was classified as an outlier (> 4.5

standard deviations above the mean) ($n = 2$). The final sample size was therefore 120 (number of females = 91; mean age = 20.7, SD = 2.5, range = 18–30).

2.1.2. Signal detection task

The task required participants to listen to bursts of noise, and to respond whether they believed that a voice stimulus was present in the noise. All participants completed two conditions: one block in which they were asked to use auditory verbal imagery whilst detecting the stimuli (the 'AVI' condition) and one block in which they were not (the 'non-AVI' condition)

Each block of the signal detection task (SDT) consisted of 80 trials, each lasting 5s (plus response time). Fig. 3.1 shows an illustration of a single trial of the task. In the AVI condition, participants were first presented with a 3–4 syllable sentence, in the centre of the screen, for 1.5s. In the non-AVI condition, participants were simply presented with a blank screen for 1.5s. This was followed by an on-screen countdown, which consisted of a shrinking circle, to mimic a '3, 2, 1...' countdown. (It was not appropriate to use a verbal on-screen countdown, as this may have interfered with processing of the presented sentence.) Pink noise (which consists of equal energy per octave; generated using Audacity 2.0.2) began playing through the provided headphones, simultaneously with the countdown. The countdown was followed by a fixation cross, which was present on the screen for 2s. The participants were informed that, if there was a voice present in the noise, it would only appear when the fixation cross was present. Participants were then prompted to respond, with a button press, as to whether they believed a voice was present in the noise.

The sentence presented to the participant always took the form of the words 'I am', followed by an adjective (i.e., the participant was required to repeat a sentence about

themselves). Half of the trials presented a positively valenced sentence (e.g. 'I am happy'), whereas the other half presented negatively valenced sentences (e.g. 'I am sad'). These sentences were rated for valence on a Likert scale (1 = very negative; 7 = very positive), by a separate subset of participants ($N = 13$), none of whom participated in the main experiments. The ratings of the words used for positive sentences ($M = 6.04$, $SD = 0.42$) were significantly higher than the words used for negative sentences ($M = 2.13$, $SD = 0.66$), ($t(65.9) = 31.53$, $p < .001$, equal variances not assumed). (Additionally, none of the mean ratings for any individual positive words were rated lower than any of the negative words.)
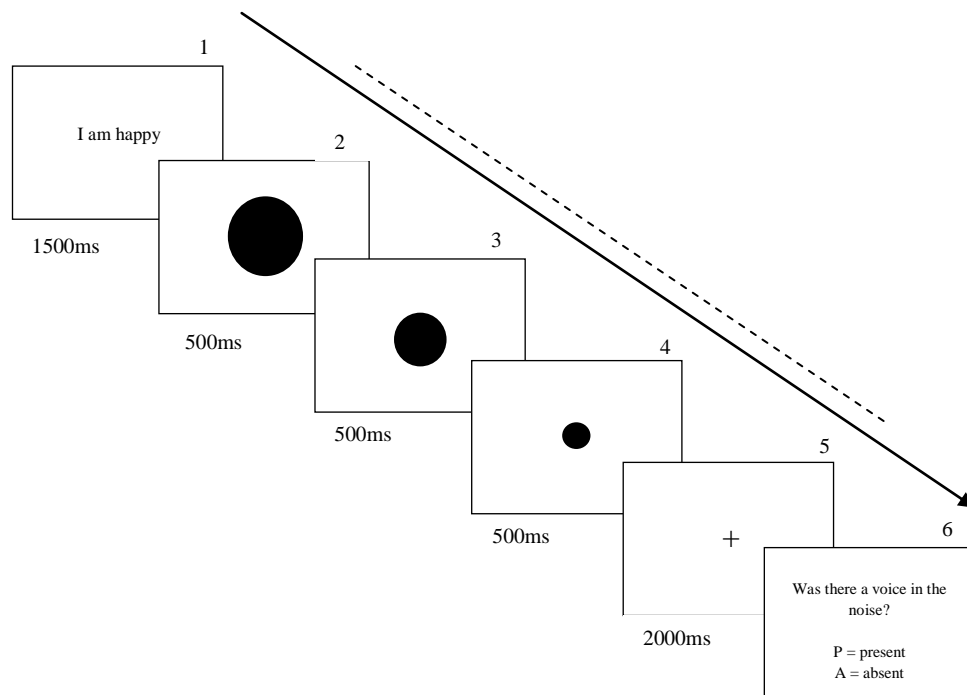
Figure 3.1: Schematic representation of a single trial in the AVI condition of the signal detection task. A sentence is presented to the participant (Screen 1), followed by a 1500ms countdown (Screens 2-4), followed by a fixation cross, which, on voice-present trials, was accompanied by a voice stimulus (Screen 5). Participants were instructed to 'imagine saying' the presented sentence when they saw the fixation cross, and then provide a response as to whether they believed a voice was present during Screen 6. The proportion of the trial during which pink noise played is indicated by the dashed line.

In 44/80 trials, a male voice stimulus was embedded in the noise as the fixation cross

appeared on the screen, lasting for 1.5 seconds ('voice-present trials'). The voice

stimuli always spoke the words previously presented to the participant on-screen on

that trial (that is, participants could specifically detect the words they had been

presented with). Within the task, the signal-to-noise ratio (voice stimulus-to-noise

volume ratio) varied between four different levels, based around the auditory

threshold of pilot participants, none of whom participated in the main experiments.

This was designed to maximise the ambiguity of the presented stimuli. In the

remaining 36/80 trials, no voice was embedded in the noise ('voice-absent trials').

The 44:36 ratio of present:absent trials was roughly based on previous auditory

signal detection studies (e.g., Barkus, Stirling, Hopkins, McKie, & Lewis, 2007),

who included more voice-present than voice-absent trials, presumably to elicit a

larger bias towards responding 'present' in all participants. Signal-to-noise ratio and

the presence/absence of a voice stimulus were balanced across valence of the

presented sentence.

2.1.3. Measure of hallucination-proneness and AVI-usage

The Revised Launay-Slade Hallucination Scale (LSHS-R) (9 items) was used to

assess hallucination-proneness. This measure was adapted by McCarthy-Jones and

Fernyhough (2011), from the longer Launay-Slade Hallucination Scale used by

Morrison, Wells, and Nothard (2000), and has previously shown high internal

reliability. The scale assesses proneness to hallucinatory experiences in both the

auditory and visual modality.

Participants were also asked to estimate the extent to which they had generated AVI

at the correct time point, giving a number between 0–100.

2.1.4. Procedure

Participants wore over-the-ear headphones (Logik LHHIFI10) to complete the task.

They were informed that they would be listening to bursts of noise, and listening out

for a voice in the noise, responding present/absent with a button press. All

participants were told that some voices would be easier to hear, whereas others

would be quieter and harder to detect, although they were not informed how often a

voice was likely to be present. For the AVI condition, participants were instructed to

"imagine saying the sentence to yourself silently", at the same time as the fixation

cross appeared on the screen. They were also informed that, if there was a voice

present in the noise, it would be presented at the same time as the fixation cross.

During the practice phase of the AVI condition, participants were asked to speak the

sentence out loud at the required time point for the first four trials, to ensure that they

understood the instructions. If the participant did not vocalise the sentence at the

appropriate time point, the practice trials were repeated until they were able to

perform the task as requested. After the practice trials, participants were asked

whether they understood the instructions relating to using AVI, and offered the

chance to repeat the practice if unsure. In the non-AVI condition, participants were

simply asked to detect a voice in the noise, but not given any instructions about

imagining a voice.

All participants completed both conditions of the SDT. The order in which they

completed the tasks was counterbalanced across participants. Between the two

blocks of trials, participants completed the self-report items (see Section 2.1.2), and

completed a variety of other task-based measures (the results of which are reported

in Chapter 4 of this thesis).

2.1.5. Data analysis

Performance on the SDT was analysed using signal detection theory. For each trial response, there were therefore four possible outcomes: hit (voice-present, 'present' response), miss (voice-present, 'absent' response), correct rejection (voice-absent, 'absent' response) and false alarm (voice-absent, 'present' response). From these, signal detection parameters relating to response bias (β) and sensitivity (d') were calculated. Following Stanislaw and Todorov (1999), β was calculated as follows:

$\beta = e\left\{\frac{Z(FA)^2 - Z(H)^2}{2}\right\}$. d' is defined as the difference between the standardised hit rate and false alarm rate.

There were two within group variables: task condition and sentence valence. A median split was also performed on the data according to score on the LSHS-R, grouping the participants into high (scoring $\geq$ 15 on the LSHS-R, $N = 61$) and low (scoring < 15, $N = 59$) hallucination-proneness; hence there was one between group variable: hallucination-proneness group (high/low). We therefore performed a $2 \times 2 \times 2$ mixed model ANOVA, with response bias (β) as the dependent variable, and task condition (AVI/non-AVI), sentence valence (positive/negative) and hallucination-proneness group (high/low) as independent variables. This analysis was also repeated with sensitivity (d') as the dependent variable, to test whether the manipulation affected participants' ability to distinguish the voice from the noise. Where data was non-normally distributed, Mann-Whitney U tests were used during further analysis.

2.2. Results

Participants generally reported being able to complete the task as instructed without difficulty, and reported using AVI with the fixation cross a relatively high amount ($M = 91.7$, $SD = 9.0$). There was no difference in the amount of AVI-usage reported between participants in the high ($M = 90.58$, $SD = 9.2$) and low ($M = 92.88$, $SD = 8.6$) hallucination-proneness groups: $t(117) = 1.40$, $p = .16$. Descriptive statistics for response bias (β) and sensitivity (d') to the auditory SDT under the two conditions (AVI/non-AVI), for positively and negatively valenced sentences, and for high and low hallucination-prone participants, are shown in Table 3.1.

2.2.1. Response bias (β)

For descriptive statistics, see Table 3.1. For β (response bias), a $2 \times 2 \times 2$ mixed model ANOVA showed a main effect of task condition (AVI/non-AVI): $F_{(1, 118)} = 5.99$, $p = .016$, showing that participants performed with a lower response bias in the AVI condition ($M = 2.41$, $SD = 2.7$) compared to when not using AVI ($M = 2.97$, $SD = 3.01$). There was no main effect of hallucination-proneness ($F_{(1, 118)} = 0.43$, $p = .51$). There was a task condition $\times$ hallucination-proneness interaction: $F_{(1, 118)} = 4.47$, $p = .037$) (see Fig. 3.2a). Further analysis using Mann-Whitney U tests showed that the effect of AVI appeared to be specific to highly hallucination-prone participants, who had a significantly lower response bias in the AVI condition ($Mdn = 1.12$) than in the non-AVI condition ($Mdn = 1.79$): $z = 3.51$, $p < .001$, $r = .45$. In the low hallucination-prone participants, there was not a significant difference in β between the AVI condition ($Mdn = 1.50$) and the non-AVI condition ($Mdn = 1.47$): $Z = 0.11$, $p = .91$, $r = .01$. There was no effect of sentence valence on β: $F_{(1, 118)} = 1.18$,

$p = .28$, nor any interactions between sentence valence and any other variables (all ps > .13).

Table 3.1: Descriptive statistics for Experiment 1, showing performance on the auditory signal detection task in the AVI and non-AVI task conditions, for high and low hallucination-prone participants, for positively and negatively valenced stimuli. (*M*, *SD*). AVI = auditory verbal imagery condition. Non-AVI = non-auditory verbal imagery condition. + = positively valenced statements. - = negatively valenced statements. β = response criterion. d' = task sensitivity.

Hallucination-proneness	Valence	AVI		Non-AVI	
		β	d'	β	d'
High	+	1.74 (1.5)	1.12 (0.6)	2.55 (1.8)	0.98 (0.6)
	-	1.83 (1.6)	1.10 (0.7)	2.47 (1.8)	1.03 (0.5)
Low	+	2.37 (1.9)	0.92 (0.6)	2.36 (1.9)	0.93 (0.6)
	-	2.15 (1.6)	0.90 (0.5)	2.17 (1.6)	1.00 (0.6)

2.2.2. Sensitivity (d')

For descriptive statistics, see Table 3.1. For d' (sensitivity), there was no effect of task condition: $F_{(1, 118)} = .148$, $p = .701$, nor any interaction between task condition and hallucination-proneness: $F_{(1, 118)} = 2.39$, $p = .125$. There was no effect of valence on d': $F_{(1, 118)} = .316$, $p = .575$, nor any interactions between valence and any other variables (all ps > .23).

2.3. Discussion

The first key finding from Experiment 1 was that when participants were instructed to use AVI during a signal detection task, there was a significant drop in response

bias. That is, participants were more likely to respond that a voice was presented in the noise when they used imagery of the same verbal stimulus, regardless of whether it was actually presented. However, using AVI did not affect participants' sensitivity (ability to distinguish between the speech and the noise). Importantly, the effect of AVI on response bias was specific to participants who scored highly on self-reported hallucination-proneness, whilst there was no difference between the imagery conditions in participants who reported few hallucinatory experiences. This supports previous findings suggesting that a bias in reality monitoring or reality discrimination may lead to auditory verbal imagery (such as inner speech) becoming misattributed to an external source. Equally, it is possible that use of AVI caused participants to exhibit a lower response bias. In this light, previous findings of lower response bias in signal detection tasks in hallucination-prone participants could be due to a higher rate of spontaneous AVI usage. This interpretation, though, is not supported by the present data, which indicated that there was no difference in AVI-usage between individuals scoring high and low in hallucination-proneness. However, since all participants were instructed to use AVI, this may have masked potential differences in spontaneous AVI usage (a possibility that is explored further in Experiment 2, below).

Contrary to our prediction, there was no effect of the emotional valence of the imagined sentence on performance, nor any interaction between valence and any other variables. This is consistent with previous research showing that, on a source memory task, the emotional valence of the stimuli did not affect performance or interact with hallucination-proneness (Bendall, et al., 2011).

3. Experiment 2

The first experiment reported here showed that using AVI during auditory signal detection led to a reduction in response bias, especially in highly hallucination-prone individuals. However, one concern is that simply cuing participants with a sentence to imagine could have altered task performance, and that the observed effect could be due to priming of the sentence, rather than the use of AVI specifically (although the observed interaction with hallucination-proneness would still be of interest). The data from Experiment 1 is not capable of addressing this concern. Because almost all participants reported a high level of AVI use, it was not possible to determine whether the effect was specific to participants who engaged in high levels of AVI. To investigate this, we therefore conducted a second experiment, using identical stimuli, in which participants were not instructed to use AVI, but were still cued with a sentence before each signal detection trial. After task completion, participants were asked to introspectively report the extent to which they felt they had used AVI whilst performing the task. The rationale for this design was that it made it possible to investigate whether signal detection performance was associated with AVI use, even when participants were not explicitly instructed to use it. This design also enabled investigation of whether hallucination-prone participants reporting using more AVI during the task (and hence had a lower response bias), or whether the amount of AVI usage was similar between groups (and any lowered response bias was solely due to an interaction between AVI usage and hallucination-proneness, as suggested in Experiment 1).

Therefore, for Experiment 2, we predicted that there would only be a difference between the two task conditions in hallucination-prone individuals who reported using high levels of AVI whilst performing the SDT.

3.1. Materials and Methods

3.1.1. Participants

The sample consisted of 60 participants from the undergraduate and staff population of Durham University, UK, none of whom had taken part in Experiment 1 (number of females = 48; mean age = 19.73, *SD* = 2.5, range = 18–30).

3.1.2. Procedure

Using identical stimuli and equipment as in Experiment 1, participants completed two blocks of the SDT. Participants were given the same instructions for how to complete the task, with the only difference being that they were given no instructions relating to AVI. Therefore, they completed two conditions: a 'non-cued' condition (identical to the non-AVI condition in Experiment 1) and a 'cued' condition (in which the same sentences as in Experiment 1 were presented before each burst of noise, but there were no AVI instructions). Participants were informed that, in the cued condition, the sentence they were presented on-screen would be the same as the voice they were instructed to detect, although they were not required to be able to comprehend the sentence in the noise to respond 'yes'.

As in the first experiment, participants completed the 9 item LSHS-R as a measure of hallucination-proneness (see Section 2.1.2) between the two blocks of the SDT. After completion of the tasks, participants were asked to what extent they felt they had used AVI during the cued condition, whilst listening for a voice, providing an

answer for the question 'This question relates to the task in which you were presented with a sentence before listening to the noise. When the fixation cross appeared on the screen, did you find yourself using 'inner speech' to say the previously presented sentence? If yes, what percentage of the time do you think you did this? (0–100)'.

3.1.3. Data analysis

As in Experiment 1, we performed a median split on the data according to LSHS-R score (high: ≥ 15, $N = 32$; low: < 15, $N = 28$). We also performed a median split on the data according to the amount of AVI reported by the participants (high: ≥ 75, $N = 32$; low: < 75, $N = 28$) and conducted a $2 \times 2 \times 2$ mixed model ANOVA, with task condition (cued/non-cued) as a within-subject variable, and hallucination-proneness (high/low) and AVI use (high/low) as between-subject variables. Due to the lack of effect of valence in Experiment 1, we did not include valence as a within-subject variable in this experiment. The dependent variable was response bias (β) on the SDT. The analysis was also repeated using sensitivity (d') as the dependent variable.

3.2. Results

Participants reported using AVI ('inner speech') a relatively high amount, considering that no instructions were given ($M = 66.83$, $SD = 29.1$), although estimates ranged from the bottom to the top of the scale (range = 0–100). A Mann-Whitney U test indicated that there was no difference between the high ($Mdn = 77.50$) and low ($Mdn = 75.00$) hallucination-proneness groups in the amount of AVI retrospectively reported ($U = 436$, $p = .86$, $r = .02$).

3.2.1. Response bias (β)

For descriptive statistics, see Table 3.2. The $2 \times 2 \times 2$ (task condition × AVI-usage × hallucination-proneness) mixed model ANOVA with *β* as the dependent variable indicated that there was no effect of task condition ($F_{(1, 56)}$ = .097, *p* = .76) on *β*. That is, the presentation of the to-be-detected sentence did not alter participants' response biases. There was no interaction between condition and hallucination-proneness ($F_{(1, 56)}$ = 2.17, *p* = .146). There was also no interaction between task condition and inner speech usage ($F_{(1, 56)}$ = .01, *p* = .91). The three-way interaction between task condition, hallucination-proneness and inner speech usage was not significant ($F_{(1, 56)}$ = 3.16, *p* = .08). However, since the interaction was close to significance, we explored the result further by conducting two 2×2 [AVI-usage × hallucination-proneness] ANOVAs, for the cued and non-cued conditions separately.

For the non-cued condition, there was no effect of hallucination-proneness ($F_{(1, 56)}$ = 1.91, *p* = .17), no effect of AVI-usage ($F_{(1, 56)}$ = 2.36, *p* = .13) and no interaction between hallucination-proneness and AVI-usage ($F_{(1, 56)}$ = .44, *p* = .51). This seems unsurprising, since the measure of AVI-usage specifically asked participants to estimate their usage of AVI only during the cued condition. For the cued condition, there was no main effect of hallucination-proneness ($F_{(1, 56)}$ = .07, *p* = .80), and there was a trend towards an effect of AVI-usage ($F_{(1, 56)}$ = 3.75, *p* = .058), with participants who reported high levels of AVI (*M* = 1.99, *SD* = 2.59) showing a lower response bias than those who reported low levels of AVI (*M* = 3.56, *SD* = 3.41).

Importantly, in the cued condition, there was a significant interaction between hallucination-proneness and AVI-usage ($F_{(1, 56)}$ = 9.12, *p* = .004). This interaction effect was explored by conducting two Mann-Whitney *U* tests for the cued

93

condition, comparing β between the high/low AVI-usage groups for both the high and low hallucination-proneness groups. This showed that, for the low hallucination-proneness group, there was no significant difference in β between the high (*Mdn* = 1.60) and low (*Mdn* = 1.80) AVI-usage groups (*U* = 93.0, *p* = .84, *r* = .04). However, for the high hallucination-proneness group, there was a significant difference in β between participants who reported high levels (*Mdn* = 1.01) of AVI-usage and those who reported low-levels (*Mdn* = 2.53) of AVI-usage (*U* = 66.0, *p* = .02, *r* = .41).

Table 3.2: Descriptive statistics for Experiment 2, showing performance on the auditory signal detection task, for high and low hallucination-prone participants, who reported high and low levels of AVI (*M*, *SD*). AVI = auditory verbal imagery; β = response bias; d' = sensitivity.

Hallucination-proneness	AVI-usage	Cued		Non-cued	
		β	d'	β	d'
High	High	1.04 (0.4)	0.87 (0.6)	1.38 (0.7)	0.86 (0.6)
	Low	4.68 (4.2)	1.40 (0.5)	3.25 (3.6)	1.14 (0.6)
Low	High	3.07 (3.5)	1.23 (0.6)	3.12 (3.7)	1.03 (0.5)
	Low	2.27 (1.6)	1.04 (0.7)	3.87 (4.4)	1.11 (0.7)
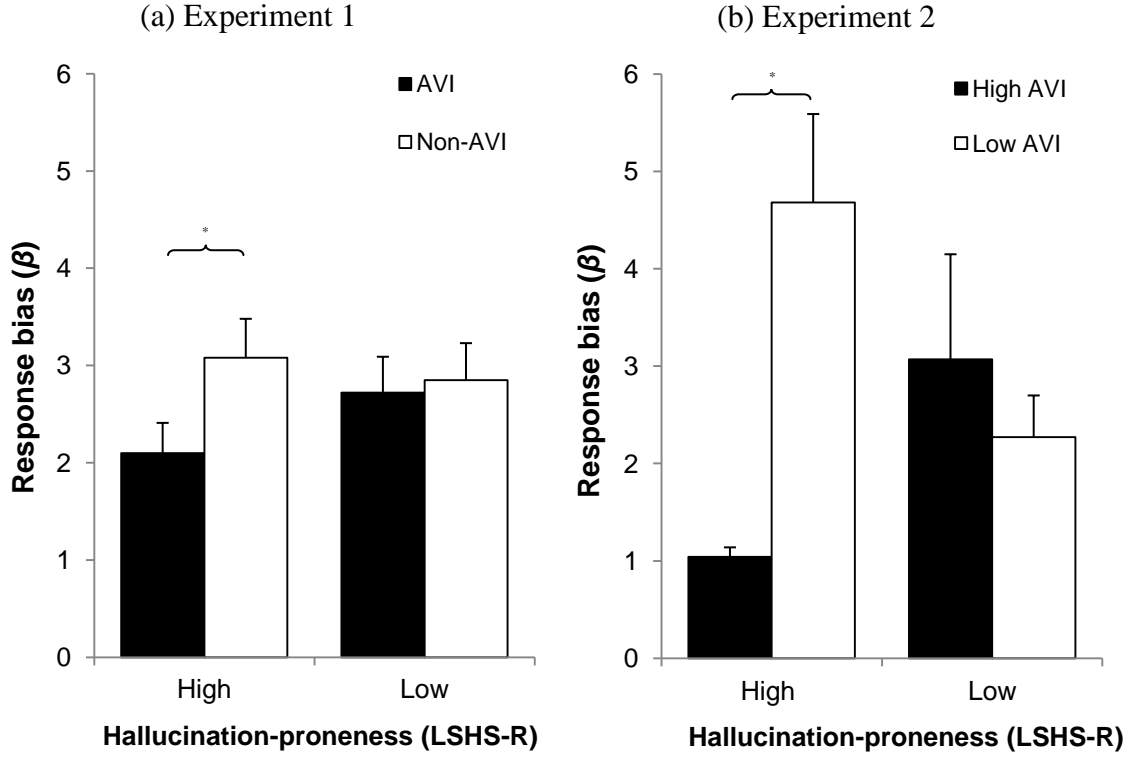
Figure 3.2: Response bias (*β*) in Experiment 1 & 2 auditory signal detection task. (a) Performance on the AVI and non-AVI condition, in high and low hallucination-prone participants. AVI = auditory verbal imagery; Non-AVI = non-auditory verbal imagery. (b) Performance on the cued condition of the SDT, split by reported level of AVI-usage, and high and low hallucination-proneness. High AVI = participants who reported levels of AVI above the median; Low AVI = participants who reported levels of AVI below the median. Error bars = 1 SEM. * *p* < .01.

3.2.2. Sensitivity (d')

For descriptive statistics, see Table 3.2. A 2 × 2 × 2 (task condition × AVI-usage × hallucination-proneness) mixed model ANOVA with *d'* as the dependent variable showed no significant effect of task condition ($F_{(1, 56)} = 1.98$, $p = .165$). As in Experiment 1, there was no interaction between task condition and hallucination-proneness ($F_{(1, 56)} = .254$, $p = .62$). There was also no interaction between task condition and AVI-usage ($F_{(1, 56)} = .02$, $p = .90$). However, as with response bias (*β*), there was an interaction, at the trend level, between task condition, hallucination-

proneness and AVI-usage ($F_{(1, 56)} = 3.74$, $p = .058$). As with β, two 2×2 [hallucination-proneness $\times$ AVI-usage] ANOVAs were conducted to explore this further. For the non-cued condition, there was no effect of hallucination-proneness ($F_{(1, 56)} = .18$, $p = .68$), no effect of AVI-usage ($F_{(1, 56)} = 1.35$, $p = .25$), nor any interaction between hallucination-proneness and AVI-usage ($F_{(1, 56)} = .36$, $p = .55$). For the cued condition, there was no effect of hallucination-proneness ($F_{(1, 56)} < .01$, $p = .99$), nor an effect of AVI-usage ($F_{(1, 56)} = 1.11$, $p = .30$). However, for the cued condition, there was an interaction between hallucination-proneness and AVI-usage ($F_{(1, 56)} = 5.23$, $p = .03$). Mann-Whitney U tests showed that, for participants in the low hallucination-proneness group, there was no difference in d' between the high ($Mdn = 1.23$) and low ($Mdn = 0.99$) AVI-usage group ($U = 78$, $p = .37$, $r = .17$). For participants in the high hallucination-proneness group, those who were in the high AVI-usage ($Mdn = 0.85$) group had a significantly lower d' score than those in the low AVI-usage group ($Mdn = 1.37$) ($U = 57$, $p = .008$, $r = 0.47$).

3.3. Discussion

Experiment 2 used identical auditory and visual stimuli as Experiment 1; only the task instructions differed, in that participants were not told that they should use AVI. After completing the task, participants estimated the extent to which they had spontaneously engaged in AVI after being cued with a sentence. The results showed that both response bias and sensitivity were affected by the presence of a sentence cue only in participants who reported high levels of hallucination-proneness *and* reported using high levels of AVI whilst detecting a voice stimulus (although it should be noted that the three-way interactions only reached trend levels of significance). Nevertheless, results seemed to indicate that if the sentence cue did not

cause participants to use high levels of AVI, it did not have an effect on task performance; however, if participants used high levels of AVI, only the participants who scored highly on self-reported hallucination-proneness showed a lowered response bias, and reduced sensitivity, when cued with a sentence.

These results are partially consistent with the results from Experiment 1: they indicate that highly hallucination-prone individuals show a lower response bias when using AVI. However, unlike Experiment 1, the results from Experiment 2 indicated that use of AVI also affected sensitivity to the task in highly hallucination-prone individuals. This was an unexpected finding, which may be explained by a greater increase in the number of 'false alarm' responses relative to the increase in 'hit' responses. That is, if the participant mistook internally-generated AVI for an externally located perception, it may have had a relatively smaller effect on the hit rate, especially if presentation of a stimulus at a low signal-to-noise ratio affected performance. For example, the presentation of voice stimuli (even below a participant's auditory threshold) may have interfered with the likelihood that internally generated AVI was mistaken as external.

It should be noted that the statistical power for the three-way interaction in Experiment 2 is low (observed $\beta = .42$). Given that exploration of the interaction was based on a finding that was above the traditional level of statistical significance, this finding should be replicated with a larger sample size.

4. General Discussion

To summarise, the two experiments reported in this paper examined the effect of the generation of auditory verbal imagery (AVI) on auditory signal detection, in participants who reported high or low levels of hallucination-proneness. Experiment 1 showed that, when instructed to use AVI, participants showed a lower response bias, being more willing to respond that a voice was present in noise, compared to performance on a standard auditory signal detection task. Further analysis showed that this effect was specific to participants who score highly on self-reported hallucination-proneness. Emotional valence of the material being imagined did not affect performance. Experiment 2 compared performance on a standard auditory signal detection task, and a variant of the task in which participants were cued with a sentence to detect, but not given any instructions to use AVI. The results suggested that hallucination-prone participants only showed a lower response bias when they retrospectively reported using AVI, despite not being instructed to do so (a finding only at trend levels of significance). In Experiment 2, counter to expectations, task sensitivity was also affected by usage of AVI.

These findings provide support for models of auditory verbal hallucinations (AVHs) which suggest that they may result from an external misattribution of an internal mental event, such as inner speech (Ditman & Kuperberg, 2005; Frith, 1992; Jones & Fernyhough, 2007b). The present studies partially support previous findings which have shown lower response biases in auditory signal detection, in both clinical and non-clinical samples that report frequent hallucinatory experiences (Brookwell et al., 2013), and extend the findings by showing that hallucination-prone individuals only showed a lower response bias when using AVI. As far as we are aware,

previous studies that have linked performance on signal detection tasks to hallucinations have not incorporated variation in AVI/inner speech usage into their study design.

Given that there was no association between level of reported AVI and level of reported hallucination-proneness, the results cannot be explained in terms of increased AVI usage in hallucination-prone individuals. Instead, the results suggest that when hallucination-prone individuals do use AVI, it is more liable to become externally misattributed. This is consistent with the previously outlined inner speech models of AVHs. The present study does not, however, provide evidence to distinguish between precise mechanisms at play in reality discrimination or reality monitoring biases. It is possible that hallucination-prone participants misattributed AVI due to high levels of vividness of the imagery, making it harder to distinguish from a 'real' perception. High levels of vividness of mental imagery may be a trait shared by hallucination-prone individuals, which could lead to a higher likelihood of external misattributions. Alternatively, low levels of cognitive effort associated with AVI generation may have led to a similar effect in this group. Neuroscientific findings describing activations in inner speech and those occurring during AVHs have implicated speech production areas, as well as primary and secondary auditory cortical regions, in the generation of AVHs (Allen, Larøi, McGuire, & Aleman, 2008), as well as showing higher levels of activity in auditory cortical (including speech perception) regions when patients with a diagnosis of schizophrenia use inner speech (Simons et al., 2010). Ford et al. (2001) previously showed that inner speech usage in hallucinating individuals was not associated with the same cortical attenuation in response to an external stimulus as in non-hallucinating individuals. In combination with the present results, this may imply that hallucination-prone

99

individuals' inner speech may be associated with higher levels of vividness, reflected in higher levels of activity in speech perception regions. This is supported by neuroimaging findings showing that auditory mental imagery (in a non-clinical sample) rated as higher in vividness is associated with higher levels of activity in speech perception regions (Zvyagintsev, Clemens, Chechko, Mathiak, Sack, & Mathiak, 2013). This is also consistent with the conclusions of Aleman et al. (2003), who interpreted the effects of auditory imagery on task performance as evidence of higher perceptual detail in the AVI of hallucinators, and may suggest that earlier findings relating to the interaction between imagery and perception (Farah & Smith, 1983) are linked to reality discrimination biases through the perceptual detail involved in auditory imagery.

Unexpectedly, hallucination-prone participants did not show a lower response bias on signal detection overall: the effect was only observed when using AVI. In this respect, the results are inconsistent with previous findings (e.g., Barkus et al., 2007; Rankin & O'Carroll, 1995; Varese, Barkus, & Bentall, 2011) which have suggested that hallucination-proneness is associated with a lower response bias in typical auditory signal detection. It is possible that task differences (for example, cueing the participants with a fixation cross at the point of the voice stimulus presentation, even in non-AVI versions of the task) could have affected performance in our study, and may make our results in non-AVI conditions non-comparable with previously conducted research. It is possible, for example, that presentation of the cross to cue voice presentation may have focused attention on voice detection, and therefore reduced the rate of spontaneous AVI, which may have masked the association with hallucination-proneness. In future experiments, it would be informative to include a condition in which no voice presentation cue is included, to test this hypothesis.

One possible objection to the interpretation of this data as relating directly to the external misattribution of internal mental imagery relates to the role of working memory, and the cognitive load associated with generating AVI during the task. Research has previously shown that increasing working memory load can lead to a reduction in the sense of agency over self-generated actions (Hon, Poh, & Soon, 2013). From our data, it is not possible to rule out the possibility that the increased working memory load (by presenting a sentence to be detected) may have interacted with hallucination-proneness, which could underlie the observed effect; however, this explanation seems unlikely, given the relatively light cognitive load involved in our task. Hon, Poh, and Soon, for example, did not find an effect of working memory on the sense of agency using a lower working memory load (two presented items), but did with a higher load (six presented items). This therefore seems like an unlikely explanation for our results.

It is also not possible to rule out that attentional processes may underlie the observed effect; for example, heightened attention to the to-be-detected stimuli in the AVI condition may have increased the participants' willingness to respond that a voice was present (although the reverse could also be the case, in that heightened attention could plausibly decrease willingness to respond a voice was present).Contemporary cognitive theories have suggested that biased attentional processes may underlie some AVHs (Hugdahl et al., 2008), and it is likely that reality discrimination biases and attentional biases are not wholly independent constructs. Future research, though, should investigate the relation between working memory, attentional biases and auditory signal detection in relation to hallucinations.

A key area of research will be to understand what causes some instances of AVI/inner speech to become misattributed, but not others. The present study found no evidence that negatively valenced words were more likely to become misattributed, which does not provide evidence for the hypothesis that negative, ego-dystonic thoughts may be externalised and experienced as a hallucination (Morrison, et al., 1995). This supports previous research using source memory tasks, which found that words associated with traumatic events were not more likely to be externally misattributed (Bendall, et al., 2011). Previous research has, however, shown that inducing negative affect in participants causes an increase in the number of external misattributions on a typical auditory signal detection paradigm (Smailes, Meins, & Fernyhough, 2014). This might imply that the content of the inner speech does not play a role in its misattribution, but instead a general state of negative affect may cause an increase in the likelihood of external misattributions. Future research should aim to examine the precise role played by affect in reality discrimination. An alternative (although not exclusive) possibility is that dialogic inner speech (that takes on the quality of a back-and-forth conversation), or inner speech that includes the voices of other people, may be more likely to be misattributed under conditions of high cognitive load or stress (Fernyhough, 2004; Jones & Fernyhough, 2007a). Further research that manipulates qualitative aspects of AVI and investigates their interaction with affective state is merited.

References

Aleman, A., Böcker, K. B. E., Hijman, R., de Haan, E. H. F., & Kahn, R. S. (2003). Cognitive basis of hallucinations in schizophrenia: role of top-down information processing. *Schizophrenia Research, 64*(2–3), 175-185.

Allen, P., Larøi, F., McGuire, P. K., & Aleman, A. (2008). The hallucinating brain: a review of structural and functional neuroimaging studies of hallucinations. *Neuroscience & Biobehavioral Reviews, 32*(1), 175-191.

Barkus, E., Smallman, R., Royle, N., Barkus, C., Lewis, S., & Rushe, T. (2011). Auditory false perceptions are mediated by psychosis risk factors. *Cognitive Neuropsychiatry, 16*(4), 289-302.

Barkus, E., Stirling, J., Hopkins, R., McKie, S., & Lewis, S. (2007). Cognitive and neural processes in non-clinical auditory hallucinations. *British Journal of Psychiatry, 191*, S76-S81.

Bendall, S., Jackson, H. J., & Hulbert, C. A. (2011). What self-generated speech is externally misattributed in psychosis? Testing three cognitive models in a first-episode sample. *Schizophrenia Research, 129*(1), 36-41.

Bentall, R. P., Baker, G. A., & Havers, S. (1991). Reality monitoring and psychotic hallucinations. *British Journal of Clinical Psychology, 30*, 213-222.

Brookwell, M. L., Bentall, R. P., & Varese, F. (2013). Externalizing biases and hallucinations in source-monitoring, self-monitoring and signal detection studies: a meta-analytic review. *Psychological Medicine, 43*(12), 2465-2475.

Ditman, T., & Kuperberg, G. R. (2005). A source-monitoring account of auditory verbal hallucinations in patients with schizophrenia. *Harvard Review of Psychiatry, 13*(5), 280-299.

Farah, M. J., & Smith, A. F. (1983). Perceptual interference and facilitation with auditory imagery. *Perception & Psychophysics, 33*(5), 475-478.

Fernyhough, C. (2004). Alien voices and inner dialogue: towards a developmental account of auditory verbal hallucinations. *New Ideas in Psychology, 22*(1), 49-68.

Ford, J. M., & Mathalon, D. H. (2005). Corollary discharge dysfunction in schizophrenia: can it explain auditory hallucinations? *International Journal of Psychophysiology, 58*(2-3), 179-189.

Ford, J. M., Mathalon, D. H., Kalba, S., Whitfield, S., Faustman, W. O., & Roth, W. T. (2001). Cortical responsiveness during inner speech in schizophrenia: an event-related potential study. *American Journal of Psychiatry, 158*(11), 1914-1916.

Frith, C. D. (1992). *The Cognitive Neuropsychology of Schizophrenia*. Hove: Lawrence Erlbaum Associates.

Frith, C. D., Blakemore, S.-J., & Wolpert, D. M. (2000). Explaining the symptoms of schizophrenia: abnormalities in the awareness of action. *Brain Research Reviews, 31*(2–3), 357-363.

Hon, N., Poh, J., & Soon, C. (2013). Preoccupied minds feel less control: sense of agency is modulated by cognitive load. *Consciousness and Cognition, 22*(2), 556-561.

Hugdahl, K., Løberg, E. M., Specht, K., Steen, V. M., van Wageningen, H., & Jorgensen, H. A. (2008). Auditory hallucinations in schizophrenia: the role of cognitive, brain structural and genetic disturbances in the left temporal lobe. *Frontiers in Human Neuroscience, 1*(6). doi: 10.3389/neuro.09.006.2007.

Johns, L. C., Kompus, K., Connell, M., Humpston, C., Lincoln, T. M., Longden, E., et al. (2014). Auditory verbal hallucinations in persons with and without a need for care. *Schizophrenia Bulletin, 40*(S4), S255-S264.

Johns, L. C., & van Os, J. (2001). The continuity of psychotic experiences in the general population. *Clinical Psychology Review, 21*(8), 1125-1141.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*(1), 3-28.

Jones, S. R., & Fernyhough, C. (2007a). Neural correlates of inner speech and auditory verbal hallucinations: a critical review and theoretical integration. *Clinical Psychology Review, 27*(2), 140-154.

Jones, S. R., & Fernyhough, C. (2007b). Thought as action: inner speech, self-monitoring, and auditory verbal hallucinations. *Consciousness and Cognition, 16*(2), 391-399.

Jones, S. R., & Fernyhough, C. (2009). Rumination, reflection, intrusive thoughts, and hallucination-proneness: towards a new model. *Behaviour Research and Therapy, 47*(1), 54-59.

Larøi, F., Van der Linden, M., & Marczewski, P. (2004). The effects of emotional salience, cognitive effort and meta-cognitive beliefs on a reality monitoring task in hallucination-prone subjects. *British Journal of Clinical Psychology, 43*, 221-233.

McCarthy-Jones, S. R., & Fernyhough, C. (2011). The varieties of inner speech: links between quality of inner speech and psychopathological variables in a sample of young adults. *Consciousness and Cognition, 20*(4), 1586-1593.

Morrison, A. P., Haddock, G., & Tarrier, N. (1995). Intrusive thoughts and auditory hallucinations: a cognitive approach. *Behavioural and Cognitive Psychotherapy, 23*(3), 265-280.

Morrison, A. P., Wells, A., & Nothard, S. (2000). Cognitive factors in predisposition to auditory and visual hallucinations. *British Journal of Clinical Psychology, 39*(1), 67-78.

Nayani, T. H., & David, A. S. (1996). The auditory hallucination: a phenomenological survey. *Psychological Medicine, 26*(1), 177-189.

Perky, C. W. (1910). An experimental study of imagination. *American Journal of Psychology, 21*(3), 422-452.

Perrone-Bertolotti, M., Rapin, L., Lachaux, M. B., & Lœvenbruck, H. (2014). What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. *Behavioural Brain Research, 261*, 220-239.

Peterson, M. J., & Graham, S. E. (1974). Visual detection and visual imagery. *Journal of Experimental Psychology, 103*(3), 509-514.

Rankin, P. M., & O'Carroll, P. J. (1995). Reality discrimination, reality monitoring and disposition towards hallucination. *British Journal of Clinical Psychology, 34*, 517-528.

Simons, C. J. P., Tracy, D. K., Sanghera, K. K., O'Daly, O., Gilleen, J., Dominguez, M., et al. (2010). Functional magnetic resonance imaging of inner speech in schizophrenia. *Biological Psychiatry, 67*(3), 232-237.

Smailes, D., Meins, E., & Fernyhough, C. (2014). The impact of negative affect on reality discrimination. *Journal of Behavior Therapy and Experimental Psychiatry, 45*(3), 389-395.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137-149.

Tien, A. Y. (1991). Distribution of hallucinations in the population. *Social Psychiatry and Psychiatric Epidemiology, 26*(6), 287-292.

Varese, F., Barkus, E., & Bentall, R. P. (2011). Dissociative and metacognitive factors in hallucination-proneness when controlling for comorbid symptoms. *Cognitive Neuropsychiatry, 16*(3), 193-217.

Varese, F., Barkus, E., & Bentall, R. P. (2012). Dissociation mediates the relationship between childhood trauma and hallucination-proneness. *Psychological Medicine, 42*(5), 1025-1036.

Woodward, T. S., Menon, M., & Whitman, J. C. (2007). Source monitoring biases and auditory hallucinations. *Cognitive Neuropsychiatry, 12*(6), 477-494.

Yoshie, M. & Haggard, P. (2013). Negative emotional outcomes attenuate sense of agency over voluntary actions. *Current Biology, 23*, 2028-2032.

Zvyagintsev, M., Clemens, B., Chechko, N., Mathiak, K. A., Sack, A. T., Mathiak, K. (2013). Brain networks underlying mental imagery of auditory and visual information. *European Journal of Neuroscience, 37*(9), 1421-1434.

Chapter 4

Cognitive predictors of hallucination-proneness in a non-clinical sample

The study in Chapter 3 showed that engaging participants in auditory verbal mental imagery during a signal detection task resulted in a lower response bias (participants were more willing to respond that they detected a voice in a burst of noise, irrespective of the actual presence of a stimulus). Furthermore, this effect appeared to be specific to participants who reported more frequent hallucinatory experiences, implying that difficulty distinguishing between internal and external perceptions is associated with proneness to hallucinations. This is consistent with the theory that hallucinations result from an external misattribution of inner speech, due to biased 1) source monitoring, 2) reality discrimination, or 3) self-monitoring processes. Problematically, despite being associated with performance on different tasks, these three terms have often been assumed to reflect a similar mechanism associated with hallucinations (i.e., the tendency to misattribute something internal as external). An important question is therefore whether these tasks are all associated with hallucination-proneness, and if so, whether they account for the same variance. Therefore, the following chapter utilised various task-based and self-report measures (including the signal detection data from the AVI condition in Chapter 3, Experiment 1) to explore whether these tasks index similar or independent cognitive constructs, using hierarchical regression analyses to investigate which, if any, of the tasks, predicted hallucination-proneness in a non-clinical sample.

Abstract

Cognitive models have proposed that hallucinations occur when an internal mental event is misattributed to an external source, due to atypical source monitoring, reality discrimination, or self-monitoring. It is unclear, however, whether task-based measures that have been used to assess these abilities tap into a common cognitive construct, or whether they require participants to engage in fundamentally different processes. Tendencies to use different types of inner speech, for example with a dialogic structure, have also previously been linked to auditory hallucinations in non-clinical samples. In the present study, a student sample completed a variety of tasks that have previously been linked to hallucination-proneness in clinical and non-clinical populations, and completed self-report measures relating to inner speech phenomenology. A hierarchical regression analysis was performed to investigate whether these variables accounted for unique or overlapping variance in self-reported hallucination-proneness. Results indicated that a tendency to use dialogic inner speech, and performance on reality discrimination tasks, independently predicted hallucination-proneness, whilst internal source monitoring did not. Unexpectedly, performance on an action self-monitoring task was positively associated with hallucination-proneness. These findings imply that several tasks which have been linked to one construct may relate to different underlying mechanisms.

1. Introduction

Contemporary research into hallucinations has examined the extent to which proneness to such experiences is distributed on a continuum in the general population (Johns & van Os, 2001), and there is now a large body of research indicating that hallucinations are reported by a substantial proportion of the population, at some point in their lives (Beavan, Read, & Cartwright, 2011). The most prominent cognitive models of hallucinations have proposed that they occur when an internal mental event is misattributed to an external source; for example, an auditory hallucination (AH) may be experienced when inner speech is misattributed (Jones & Fernyhough, 2007), or a visual hallucination when visual imagery is misattributed (Brebion, Ohlsen, Pilowsky, & David, 2008). Therefore, one fruitful approach has been to examine associations between self-reported 'hallucination-proneness', and performance on cognitive tasks designed to assess participants' ability to distinguish between internally and externally generated perceptions (Ditman & Kuperberg, 2005), with the hypothesis that individuals who report more hallucinatory experiences will perform differently on these tasks. These cognitive tasks have been linked to two separate theoretical strands of research: reality monitoring, embedded in the source monitoring framework (Johnson, Hashtroudi, & Lindsay, 1993), and self-monitoring embedded in motor theories of agency (Frith, 1992).

The source monitoring framework attempts to explain how we make judgements about the origin (source) of remembered information, using various characteristics such as perceptual, semantic or affective detail, or cognitive effort (Johnson et al., 1993). The term 'source monitoring', as a whole, can be divided into three categories: internal source monitoring (distinguishing between multiple internal

sources; for example, whether a word was spoken aloud or imagined), external source monitoring (distinguishing between external sources; for example, whether a word was spoken by one person or another person), or reality monitoring (distinguishing between internal and external sources; for example, whether a word was spoken by oneself, or someone else). Some theories argue that hallucinations may be instances in which reality monitoring has either failed, or become biased towards labelling information as external (Bentall, Baker, & Havers, 1991), and this is supported by numerous studies showing that hallucinating psychotic patients are more likely to incorrectly recall a word as externally generated than non-hallucinating patients (Bentall et al., 1991; Stephane, Kuskowski, McClannahan, Surerus, & Nelson, 2010; Woodward, Menon, & Whitman, 2007), especially under conditions of high cognitive effort (Bentall et al., 1991). This finding also appears to hold for non-clinical samples who report more frequent hallucinatory experiences (Larøi, Van der Linden, & Marczewski, 2004).

Some studies have, however, also indicated that a similar bias may be present for internal source monitoring tasks, towards labelling imagined items as performed/spoken aloud (Franck, Rouby, Daprati, Daléry, Marie-Cardine, & Georgieff, 2000; Gawęda, Woodward, Moritz, & Kokoszka, 2013). One possible explanation for these findings is that biases in recalling what Johnson et al. (1993) call "the actual (public) vs. imaginal (private) status of the information" (p. 3) are linked to hallucinations, and that this bias is picked up on by both internal source monitoring and reality monitoring tasks (with words spoken aloud in internal source monitoring tasks being more 'public' than those that are imagined). A recent meta-analysis showed that while reality monitoring was associated with proneness to hallucinations in non-clinical samples, there was insufficient evidence relating

internal source monitoring to hallucinations to reach any concrete conclusions (Brookwell, Bentall, & Varese, 2013).

Other strands of research relating to reality monitoring have related biases in auditory signal detection (most commonly, detecting a voice in noise) to the tendency to experience hallucinations. Unlike source memory paradigms, this requires the participant to make an 'online' decision about whether a perception is internally or externally generated, and has previously been termed 'reality discrimination' to distinguish it from the 'offline' decisions made during source memory tasks (that is, when there is a delay between stimulus presentation and the decision regarding whether it was self-generated or not) (Bentall et al., 1991). Reality discrimination has, though, been theoretically linked to source monitoring processes (Aleman, Böcker, Hijman, de Haan, & Kahn, 2003), on the grounds that such tasks require participants to decide whether any perceived voices were present in the noise, or were simply generated internally. Numerous studies have associated proneness to hallucinations with a bias in auditory signal detection towards reporting hearing voices in noise, in both clinical (Varese, Barkus, & Bentall, 2012) and non-clinical (Barkus et al., 2011) samples. Importantly, most studies show no between-group difference in *sensitivity* (the ability to distinguish between speech and noise) to the signal detection task (though see Vercammen, de Haan, and Aleman, 2008, for an exception).

Some studies have explicitly engaged participants in mental imagery whilst they perform signal detection, to investigate the role of imagery generation on reality discrimination and false perception (Aleman, et al., 2003, also see Chapter 3 of this thesis), with results indicating that hallucination-proneness was linked to a tendency to mistake self-generated imagery for an external stimulus. Various other tasks have

also been used to elicit imaginary verbal experiences, and therefore linked to source monitoring; for example, the jumbled speech task asks participants to listen to speech-like sounds and requires them to report any words they believe were present (Fernyhough, Bland, Meins, & Coltheart, 2007). However, it is unclear to what extent these tasks tap into the same cognitive mechanisms as the aforementioned source memory tasks, or whether atypical jumbled speech or signal detection performance may independently predict proneness to hallucinations. An important area of research would therefore be to investigate whether source monitoring and reality discrimination tasks account for the same variation in hallucination-proneness, or whether they appear to be measuring separate cognitive constructs.

A separate, though related, strand of research is grounded in motor theories which attempt to explain how agency is experienced for self-generated actions (self-monitoring). The most prominent theory has suggested that, when a motor command for an action is generated, the predicted consequences of that action (sent via an 'efference copy') are compared with the resulting incoming sensory information. If the prediction and incoming sensory information match, the cortical response is attenuated, and the action is experienced as self-generated. Various theories have linked aberrant self-monitoring (efference copy) mechanisms to auditory verbal hallucinations specifically[3] (Ford, Roach, Faustman, & Mathalon, 2007; Heinks-Maldonado et al., 2007), or schizophrenia more generally (Frith, 1992; Whitford, Ford, Mathalon, Kubicki, & Shenton, 2012). It is now a fairly well-replicated finding that patients with a diagnosis of schizophrenia perform poorly on tasks requiring

---

[3] It should be noted, however, that many studies, including those cited here, have not shown that efference copy mechanisms are dysfunctional in only hallucinating individuals. Instead, the studies usually show a difference in the N1 ERP response (thought to reflect an efference copy mechanism) to self-produced vocalisations, between patients with a diagnosis of schizophrenia, and healthy controls.

them to monitor their own actions without visual feedback (Frith & Done, 1989) or with distorted visual feedback (Franck et al., 2001), and it is therefore possible that an aberrant self-monitoring system may be able to explain various psychotic symptoms.

Such tasks often require participants to take control of a visually presented stimulus using a computer mouse or joystick, and, in the absence of visual feedback of the movement, discriminate which stimuli are related to movements initiated by the participant and which are not (Franck, et al., 2001; Williams & Happé, 2009). Theoretically, an aberrant efference copy system should lead to lower performance on these tasks: because participants cannot rely on external feedback, they must rely on internally generated signals to monitor their performance. Similarly to source monitoring accounts outlined above, self-monitoring theories may specifically be able to explain how internal mental events, such as inner speech, could become misattributed to an external source, and therefore experienced as a hallucination (Jones & Fernyhough, 2007). There have, however, been comparatively few studies investigating self-monitoring and hallucination-proneness in non-clinical samples, and so a fruitful area of investigation would be to study to what extent clinical findings are replicable in non-clinical samples.

A third variable that has been linked to proneness to hallucinations is phenomenology of inner speech (McCarthy-Jones & Fernyhough, 2011). A commonly cited problem with theories that explain auditory verbal hallucinations as misattributed inner speech is that heard voices tend to be experienced in the second or third person ('you', 'he/she') and in someone else's voice (Nayani & David, 1996), which does not seem consistent with a folk conceptualisation of the experience of inner speech. However, from a Vygotskian perspective, it has been

argued that typical inner speech may be dialogic in nature (i.e., involving the back-and-forth of a conversation), and may frequently include the voices of other people, representing the developmental endpoint of a gradual internalisation of dialogues held externally (Fernyhough, 1996, 2004). This would be consistent with a view of inner speech as important in self-evaluation (Morin & Michaud, 2007), which could explain why many AHs take the form of a voice commenting or instructing the individual.

A recent investigation of individual's self-reported experience of inner speech showed variations on a number of different dimensions, including a tendency to engage in dialogic inner speech, to use inner speech involving the voice of another person, to use inner speech for evaluation of one's own behaviour, and to use inner speech with a 'condensed' (abbreviated or shortened) format (McCarthy-Jones & Fernyhough, 2011). Importantly, the first three of these dimensions correlated significantly with self-reported proneness to hallucinatory experiences, and when controlling for other factors such as anxiety and depression, tendency to use dialogic inner speech was a significant predictor of proneness to AHs specifically. A subsequent study replicated the factor structure and reliability of these dimensions of inner speech, but did not replicate the unique variance in hallucination-proneness accounted for by reported dialogicality of inner speech (Alderson-Day et al., 2014), although this may have been due to the inclusion of different covariates relating to anxiety and depression.

Table 4.1: Summary of the types of monitoring often associated with AVHs, their definitions, and associated tasks. The definitions are not mutually exclusive; it is not clear to what extent the various types of monitoring refer to the same or different cognitive mechanisms.

Type of monitoring	Definition	Associated tasks
Self-monitoring	The ability to tell whether an action is self-produced or not; often associated with a sense of authorship over actions. Self-monitoring theories have traditionally been based in theories suggesting that we experience agency over our actions as a result of efference copies sent by the motor system to sensory cortices.	Error monitoring (participant must detect when a self-controlled stimulus does not concord with their movement). Action monitoring (participant must distinguish self-controlled stimulus movements from non-self-controlled movements, compared to when someone else controls the stimulus).
Source monitoring:		
Internal source monitoring	Distinguishing between two self-produced (internal) sources of information.	Source memory (did I say or imagine that word?)
External source monitoring	Distinguishing between two non-self-produced (external) sources of information.	Source memory (did that person say that word, or another person?)
Reality monitoring	Distinguishing between self-produced and non-self-produced (internal and external) sources of information.	Source memory (did I say/imagine that word, or did someone else?)
Reality discrimination	Distinguishing between self-produced and non-self-produced sources of information. This term has generally been used to refer to instances in which the participant is required to make an online (immediate) decision about the source of information.	Auditory signal detection task (detecting a voice in noise) Jumbled speech task (hearing words in jumbled speech)

An interesting question is therefore whether the tendency to engage in specific forms of inner speech can predict proneness to hallucinatory experiences independently of source memory and self-monitoring biases/deficits, or instead whether they might account for the same variance. Likewise, it is important to establish whether the assorted tasks which have been related to source monitoring or self-monitoring are all associated with proneness to hallucinations, and if so, whether they account for the same or different variance.

The present study set out to test this by asking participants to complete four tasks that have previously been used to assess source monitoring, reality discrimination or self-monitoring (in both verbal and non-verbal modalities): 1) a signal detection (reality discrimination) task that required participants to indicate the presence or absence of speech stimuli in noise, whilst using auditory imagery (using the same data as described in Chapter 3); 2) a source memory task that required participants to imagine or say words under conditions of high or low cognitive effort, and later recall how they had produced the word (internal source monitoring); 3) an 'action monitoring' task that required participants to map self-generated and non-self-generated hand movements onto a visually presented stimulus (self-monitoring); 4) an auditory task which asked participants to pick out (non-existent) verbal stimuli in 'jumbled speech'. Participants also completed self-report measures relating to inner speech phenomenology, as well as depression and anxiety. Measures of depression and anxiety were taken to test whether any observed associations were specific to hallucinations, as opposed to 1) reflecting a general style of responding on self-report measures (i.e., common method variance), or 2) reflecting an association with a general alteration in cognitive state which may be shared between hallucination-proneness and depression/anxiety.

The aim of the study was to investigate whether these tasks and self-report measures predicted hallucination-proneness (also assessed with a self-report questionnaire), and if so, whether they accounted for overlapping or unique variation in tendency to experience hallucinations.

2. Materials and Methods

2.1. Participants

120 participants were recruited from the staff and student population of a UK university, via email invitation and advertisements in various departments. 10 participants were not included in the subsequent analysis, as they did not complete all tasks due to technical failures or early withdrawal from the study. We recruited participants between the ages of 18–30 ($M = 20.79$, $SD = 2.56$, no. females = 87), as the nature of signal detection analysis required that volume levels of the auditory stimuli were kept constant across all participants, and age may have affected the sensitivity to the auditory tasks. We also specified that participants should not have any diagnosed hearing problems, and that they spoke English as a first language. All gave written informed consent, and ethical approval was given by Durham University Ethics Committee.

2.2. Self-report measures

2.2.1. Revised Launay-Slade Hallucination Scale (LSHS-R)

The LSHS-R is a 9-item questionnaire which assesses predisposition to hallucinatory experiences. It was adapted by McCarthy-Jones and Fernyhough (2011) from Morrison, Wells, and Nothard's (2000) Revised Launay-Slade Hallucination Scale,

116

which had relatively low internal reliability. The questionnaire consists of two subscales, relating to auditory hallucinations (LSHS-R$_{aud}$) (e.g., 'I have had the experience of hearing a person's voice and then found that no-one was there') and visual hallucinations (LSHS-R$_{vis}$) (e.g., 'I see shadows and shapes when nothing is there'). Each item is scored on a four-point Likert scale from 1 ('never') to 4 ('almost always'), summing to total scores from 9-36.

2.2.2. Hospital Anxiety and Depression Scale (HADS)

The HADS is a 14-item questionnaire consisting of two subscales, measuring anxiety (HADS$_{anx}$) (seven items, e.g., 'In the past month, I have felt tense and wound up') and depression (HADS$_{dep}$) (seven items, e.g., 'In the past month, I have felt 'slowed down''). It has previously been shown to have satisfactory psychometric properties (Zigmond & Snaith, 1983). Each item is scored between 0–3, summing to a total score between 0–21 for each subscale.

2.2.3. Varieties of Inner Speech Questionnaire (VISQ)

The VISQ is an 18-item questionnaire, designed to assess the self-reported phenomenological properties of inner speech. It consists of four subscales, relating to tendency to engage in dialogic inner speech (VISQ$_{dialogic}$, 4 items), tendency to use inner speech consisting of voices other than one's own (VISQ$_{other}$, 5 items), tendency to use inner speech to evaluate and motivate behaviour (VISQ$_{eval}$, 4 items) and tendency to use condensed inner speech (VISQ$_{condensed}$, 5 items). It has satisfactory psychometric properties (Alderson-Day et al., 2014; McCarthy-Jones & Fernyhough, 2011). Each item is scored between 1 ('certainly does not apply to me') and 6 ('certainly applies to me').

2.3. Task-based measures

2.3.1. Signal detection task (SDT)

In this task, participants listened to bursts of pink noise generated using Audacity 2.0.2, through over-the-ear headphones (Logik LHHIFI10), and were asked to respond with a button press (yes/no) as to whether they believed a voice to be present in the noise. At the same time, they were asked to engage in auditory imagery of the same sentence, which had previously been presented visually. The aim of the task was therefore to require the participant to decide whether any perceived sentence was self-generated (i.e., they imagined it) or non-self-generated (i.e., it appeared in the noise).

There were 80 trials, with each burst of noise lasting 3.5 seconds. In 44/80 trials, a male voice was present ('voice-present trials') at one of four possible volumes, based slightly above and below the threshold at which 50% of participants in a small pilot study tended to detect the voice. In the remaining 36 trials only noise was played ('voice-absent trials'). Voice-present and voice-absent trials were presented in a random order. Before each trial, a short sentence (3–4 syllables) was presented in the centre of the computer screen, which the participant was asked to 'imagine saying' when the fixation cross appeared on the screen. This sentence was always the same as that spoken by the voice stimulus, embedded in the noise. In each trial, participants were then cued with a countdown lasting 1.5 seconds, which started at the same time as the noise. The countdown consisted of a circle positioned in the centre of the computer screen shrinking, which was followed by the fixation cross in the centre of the screen. At the fixation cross, the voice stimulus was presented in the noise (on the voice-present trials), and the participant was asked to simultaneously

imagine saying the previously presented sentence. After each burst of noise, participants were visually cued to respond with the question 'Was there a voice present in the noise?', and pressed a button on the computer keyboard to give their response. Participants were informed that, if a voice was present, it would appear in the pink noise at the same time as the fixation cross. They were also told that voices may be present at a variety of volumes, although they were not told on how many trials a voice would be present. Before completing the main task, participants conducted a practice task consisting of 8 bursts of noise. On half of these trials, participants were asked to speak the cued sentence aloud with the fixation cross, to show that they understood the timing of the task. On the remaining trials, participants used auditory imagery as in the main task; the experimenter then checked that the participant understood the instructions regarding the auditory imagery, before the main task started.

Overall, the SDT task lasted approximately 9 minutes, including an enforced 30-second break midway through the task. The dependent variables in this task were signal detection measures: β, a measure of response bias (with a lower value indicating a lower criterion for responding that a voice was present) and d', a measure of task sensitivity (with a lower value indicating less ability to distinguish between signal and noise).

Since our previous finding demonstrated that auditory signal detection was only associated with hallucination-proneness when participants were concurrently engaged in auditory verbal imagery (see chapter 3), only this version of the task was entered into the regression analysis (whereas the signal detection performance with no auditory imagery was not included).

2.3.3. Internal source monitoring task (SMT)

This task assessed source memory for internally generated items (imagined, or spoken aloud), and was based loosely on that used by Franck et al. (2000). There were two stages to the SMT. In the word completion stage, participants were presented with a series of easy 'word pairs' (for example, 'gold and silver'), some of which they were required to say out loud, and some of which they were instructed to say to themselves using inner speech. 80 word pairs were included in the task, 40 of which were presented to the participant fully (e.g., 'black and white'), and so only requiring low cognitive effort from the participant; and 40 of which were only partially completed (e.g., 'black and w____'), requiring higher effort from the participant to complete. For each trial, the instruction ('Out Loud' or 'Inner Speech') appeared on the screen for 1250ms, followed by the word pair, for 3250ms, followed by an intertrial interval of 750ms. Participants were instructed to say the full word pair, in the manner instructed, as soon as possible after the presentation. If they did not know the correct answer to the uncompleted word pair, they were asked to indicate this with a button press. They were not informed that a memory test would follow the word completion stage. After the word completion stage, participants took a break from the task, in which they completed the jumbled speech task (see Section 2.3.4), and began completing the self-report measures (described in Section 2.2).

After 15 minutes, participants were asked to complete the recall stage of the task, in which they were presented with the second part of each word pair (e.g., 'white'), in a random order. Participants were instructed to try to recall whether they said each word 'out loud' or using 'inner speech', responding with a button press. Each word was presented in the centre of the computer screen until a response was made. The dependent variables for this task were the signal detection measures: β (response

bias) (with a lower value corresponding to a lower criterion for responding that a word was spoken aloud) and d' (sensitivity) (with a lower value indicating less ability to distinguish which words were spoken aloud and which using inner speech).

2.3.4. Jumbled speech task (JST)

This task was adapted from the stimuli used by Fernyhough, Bland, Meins, and Coltheart (2007). Participants were presented with a recording of a 'jumbled' female voice, and asked to type any words or phrases heard in the speech. The stimuli had been prepared by segmenting the speech at random silent intervals, reversed and then assembled into a continuous stream of jumbled speech. Participants listened to 12 tracks of jumbled speech (as well as one practice track), each lasting nine seconds. After each track, they were invited to type in any words or phrases heard in the speech. The dependent variable for this task was the number of syllables each participant reported hearing in all trials. Following the procedure of Fernyhough et al. (2007), words that were reported by > 10% of participants in each trial were removed from subsequent analysis as an illusion (i.e., specific words that were reported commonly may represent perception of parts of the jumbled speech that, by coincidence, sounded like real words).

2.3.5. Action monitoring task (AMT)

This task was adapted from the task used by Williams and Happé (2009), in which participants were given control of a computer mouse, and asked to determine which of a group of squares displayed on the computer screen they were able to control the movement of using the mouse (see Fig. 4.1). Whilst all squares on the screen stopped or started moving simultaneously, only one square moved in a direction consistent with the mouse movement. Participants' score corresponded to the number of

squares that were correctly identified as controlled by the mouse. There were two

conditions to the task: 1) the participant freely controlled the mouse; 2) the

participant rested their hand on the mouse, and attempted to determine which square

was controlled by the experimenter, as the experimenter moved the mouse. In both

conditions, the participant's arm was occluded from view underneath a box. In the

first, 'self-controlled' condition, the participant must therefore map the movement of

their hand onto a square on the screen, using both information derived from internal

motor processes and proprioceptive feedback from their hand. In the second, 'other-

controlled' condition, the participant could only rely on proprioceptive feedback,

since the experimenter was in control of the mouse. Therefore, following the

procedure of Williams and Happé (2009), the ability to monitor self-produced

actions using only information from internal motor processes can be derived by

subtracting the score from the other-controlled condition from the self-controlled

condition.

Table 4.2: Stimulus properties of the action monitoring task. No. squares =
number of squares presented on the computer screen, from which the participant had
to identify the target. Distractor movement (°) = the extent to which the movement of
the distractor squares could vary, relative to the movement of the computer mouse.

Level	No. squares	Distractor movement (°)
1	9	360
2	16	360
3	25	360
4	25	180
5	36	180
6	36	90

The task was designed to increase in difficulty by manipulation of both the number

of distractor squares and the degree to which the distractor squares matched the

movement of the target square (see Table 4.2 for details of the increasing difficulty

of the task). The difficulty of the task was increased after every 8 trials, and,

similarly to the methodology of Williams and Happé (2009), if participants made ≤ 2

correct responses in those trials, the condition was discontinued. This was to prevent

participants advancing in the task by chance. The dependent variable in this task was

therefore the number of squares correctly identified in the self-controlled condition,

minus the number of squares correctly identified in the experimenter-controlled

condition.
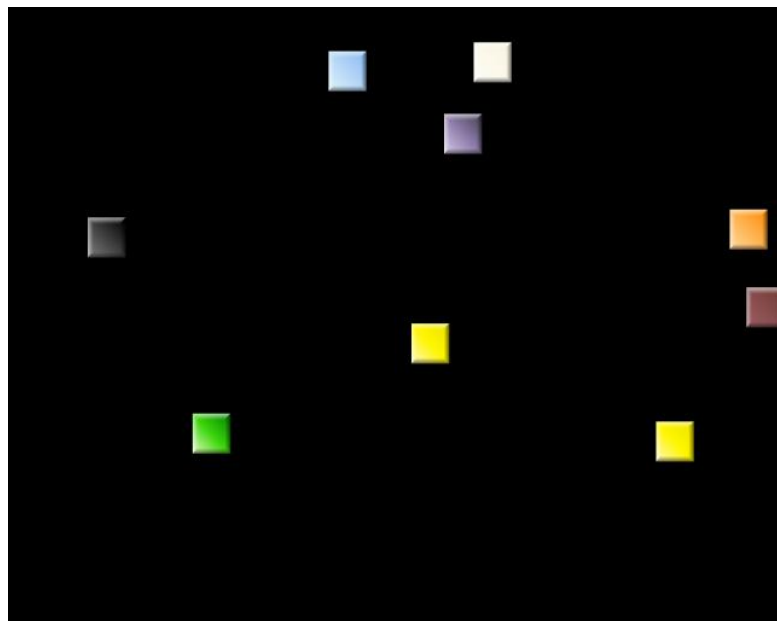
Figure 4.1: Screenshot of action monitoring task. (See Section 2.3.5 for description of task.)

3. Results

3.1. Summary

Bivariate correlations between the self-report scales and LSHS-R scores, as well as between the task-based measures and LSHS-R scores, are presented in Tables 4.3 and 4.4. After applying Bonferroni's correction for multiple comparisons, total LSHS-R score correlated significantly with the $VISQ_{dialogic}$ and $VISQ_{other}$, but not with the $VISQ_{eval}$ subscale, or with either of the HADS subscales. The only task-based measure to significantly correlate with total LSHS-R score, as well as the auditory and visual subscales, was the jumbled speech task, although these correlations were not significant after Bonferroni's correction for multiple comparisons was applied.

Table 4.3: Bivariate correlations (Spearman's) among self-report scales.

	1	2	3	4	5	6	7	8
1. LSHS-R$_{total}$	1	-	-	.406**	.393**	.266*	.242*	.166
2. LSHS-R$_{aud}$		1	.503**	.396**	.376**	.243*	.218*	.106
3. LSHS-R$_{vis}$			1	.323**	.357**	.279*	.251*	.271*
4. VISQ$_{dialogic}$				1	.477**	.481**	.294**	.147
5. VISQ$_{other}$					1	.341**	.245*	.119
6. VISQ$_{eval}$						1	.279*	.227*
7. HADS$_{anx}$							1	.514**
8. HADS$_{dep}$								1

*$p < .05$. ** $p < .0019$ (i.e. = .05/26).

The action monitoring task was significantly positively correlated with scores on the LSHS-R$_{aud}$ subscale, and the signal detection task was negatively correlated with scores on the LSHS-R$_{vis}$ subscale (i.e., a lower criterion for accepting that a signal was present corresponded to a higher score on the LSHS-R$_{vis}$) although, again, these were not significant after Bonferroni's correction was applied. Neither the high or low cognitive effort conditions of the source memory task correlated with LSHS-R scores, and are therefore treated as one SMT score (combined) in the regression analysis.

To investigate which variables may be able to predict proneness to hallucinations, we carried out a hierarchical regression analysis to investigate whether self-reported phenomenology of inner speech (VISQ) and performance on tasks purported to measure internal source monitoring, reality discrimination, and self-monitoring could predict the total score on the LSHS-R, when controlling for gender, age, anxiety and depression. We then constructed a model using the variables implicated in the hierarchical regression, using a stepwise procedure.

Table 4.4: Bivariate correlations (Spearman's) among task-based measures and hallucination-proneness measures. Correlations are given for both the high and low effort conditions of the source memory task.

	Source memory, high (β)	Source memory, low (β)	Signal detection (β)	Jumbled speech	Action monitoring
LSHS-R$_{total}$	.048	.024	-.182	.229*	.133
LSHS-R$_{aud}$	.040	.015	-.113	.193*	.194*
LSHS-R$_{vis}$	.050	-.026	-.235*	.237*	-.039

* $p < .05$. No correlations were significant at $p < .0033$ (i.e., = .05/15).

3.2. Multiple linear regression

3.2.1. Total hallucination-proneness

A hierarchical multiple linear regression (MLR) was performed, with total hallucination-proneness score (LSHS-R$_{total}$) as the dependent variable. There was no evidence of multicollinearity in the data (tolerance for all variables >.47), and Durbin-Watson scores were satisfactory. There was also no evidence that any single case exerted a significant influence on the model (all Cook's distances < .18), and inspection of the residuals indicated that the assumption of homogeneity of variance of residuals was not broken. Finally, the residuals were normally distributed ($D =$ .04, $p = .20$).

In the first step, age, gender, and self-report measures of anxiety and depression (HADS$_{anx}$ and HADS$_{dep}$) were included as independent variables. This model was significant ($F_{(4, 109)} = 4.60$, $p = .002$, $R^2 = .149$), with both gender and anxiety emerging as significant predictors of hallucination-proneness (see Table 4.5). HADS$_{anx}$ score was positively associated with hallucination-proneness, and males tended to score higher than females on the LSHS-R. In the second step, three subscales of inner speech phenomenology (VISQ$_{dialogic}$, VISQ$_{other}$, and VISQ$_{eval}$) which have previously been associated with hallucination-proneness in a non-clinical sample were included. Addition of these variables significantly improved the model ($F_{(3, 102)} = 8.37$, $p < .001$, $\Delta R^2 = .168$), with both the VISQ$_{dialogic}$ and VISQ$_{other}$ subscales significantly contributing to the model. Finally, in the third step, task-based measures hypothesized to be associated with hallucination-proneness (SMT, SDT, JST, AMT) were included. Again, the model was significantly improved ($F_{(4, 98)} = 6.47$, $p < .001$, $\Delta R^2 = .143$). Of the task-based measures, performance on the

signal detection task, jumbled speech task and action monitoring task all significantly predicted LSHS-R$_{total}$ score, although performance on the internal source monitoring task did not. In this third model, score on the VISQ$_{other}$ subscale had dropped to a trend level of significance ($p = .077$), but gender, anxiety, and VISQ$_{dialogic}$ still significantly contributed to the model. Unexpectedly, performance on the action monitoring task was positively associated with LSHS-R$_{total}$ score, indicating that individuals with higher hallucination-proneness scores performed better on this task (in contrast to our original hypothesis).

To test the specificity of the individual task-based measures, each of the four variables (SDT, JST, SMT, AMT) were entered separately into a block, in separate regression models, again controlling for age, gender, HADS$_{anx}$, HADS$_{dep}$, and the VISQ scales. This did not alter any of the main results. (That is, JST, SDT and AMT still significantly predicted LSHS-R, but SMT did not.)

The variables that significantly contributed to the above model were retained for a subsequent stepwise procedure. In this model, the following variables were entered: gender, age, HADS$_{anx}$, VISQ$_{dialogic}$, the SDT, JST and AMT (see Table 4.6). All variables significantly contributed to the model, with the exception of age, which was excluded from the model. The final model significantly predicted total hallucination-proneness ($F_{(7, 109)} = 10.32$, $p < .001$), and accounted for 41.5% of variance in LSHS-R$_{total}$ score.

Table 4.5: Summary of hierarchical multiple linear regression analyses for variables predicting hallucination-proneness ($N = 110$)

Variable	B	SE B	β
Step 1			
Constant	22.05	3.38	
Gender	-2.63	0.86	-0.28**
Age	-0.21	0.14	-0.14
HADS Anxiety	0.27	0.12	0.28*
HADS Depression	0.01	0.16	0.01
Step 2			
Constant	16.21	3.30	
Gender	-1.96	0.81	-0.21*
Age	-0.24	0.13	-0.16
HADS Anxiety	0.12	0.11	0.12
HADS Depression	0.02	0.15	0.02
VISQ Dialogic	0.16	0.07	0.21*
VISQ Other	0.15	0.06	0.23*
VISQ Evaluative	0.12	0.09	0.13
Step 3			
Constant	14.57	3.07	
Gender	-1.85	0.78	-0.20*
Age	-0.27	0.12	-0.18*
HADS Anxiety	0.21	0.1	0.22*
HADS Depression	-0.04	0.14	-0.03
VISQ Dialogic	0.21	0.07	0.28**
VISQ Other	0.11	0.06	0.16
VISQ Evaluative	0.14	0.08	0.15
SMT (response bias)	-0.27	0.25	-0.09
IS-SDT (response bias)	-0.28	0.12	-0.19*
JST (syllables)	0.07	0.02	0.25**
AMT (self – other)	0.10	0.04	0.23**

$* p < .05. ** p < .01. *** p < .001.$

Table 4.6: Summary of stepwise regression analysis for variables that significantly contributed to hierarchical regression predicting total hallucination-proneness

Variable	B	SE B	β
Constant	11.46	1.87	
VISQ Dialogic	0.29	0.06	0.39***
IS-SDT (response bias)	-0.32	0.12	-0.21*
Gender	-2.10	0.74	0.22**
JST (syllables)	0.07	0.02	0.26**
AMT (self – other)	0.09	0.04	0.19*
HADS Anxiety	0.19	0.08	0.20*

*$p < .05$. ** $p < .01$. *** $p < .001$.

3.2.2. Auditory and visual hallucination-proneness

To investigate whether these variables might specifically be related with auditory or visual hallucination-proneness, a hierarchical regression analyses with both individual subscales of the LSHS-R (LSHS-R$_{aud}$ and LSHS-R$_{vis}$) as independent variables was conducted, and the remaining LSHS-R subscale was included as a predictor. In this way, we investigated whether any inner speech variables or task-based measures were related to a certain modality of hallucination, when controlling for the other modality.

With LSHS-R$_{aud}$ as the dependent variable, the same variables were entered in the same three steps as when predicting LSHS-R$_{total}$, as well as LSHS-R$_{vis}$ in the first step. In the first step, gender, age, LSHS-R$_{vis}$, HADS$_{anx}$, and HADS$_{dep}$ were included in the model, which significantly predicted LSHS-R$_{aud}$ ($F_{(5, 104)} = 10.65$, $p < .001$, $R^2 = .339$), although only LSHS-R$_{vis}$ significantly predicted LSHS-R$_{aud}$ ($\beta = .515$, $p <$

.001). In the second step, the three inner speech subscales (VISQ$_{dialogic}$, VISQ$_{other}$, VISQ$_{eval}$) were added, which again resulted in a significant improvement to the model ($F_{(3, 101)} = 3.03$, $p = .033$, $\Delta R^2 = .055$), although none of the individual inner speech variables significantly predicted LSHS-R$_{aud}$. (VISQ$_{dialogic}$ only approached significance; $\beta = .172$, $p = .072$.) In the third step, the four task-based measures were included, which again significantly improved the model ($F_{(4, 97)} = 3.54$, $p = .010$, $\Delta R^2 = .077$), with LSHS-R$_{vis}$ ($\beta = .346$, $p < .001$), age ($\beta = -.173$, $p = .030$), VISQ$_{dialogic}$ ($\beta = .219$, $p = .022$), and the AMT ($\beta = .267$, $p < .001$) significantly predicting LSHS-R$_{aud}$. Therefore, in the final model, when controlling for visual hallucination-proneness, the variables that significantly predicted auditory hallucination-proneness were age, VISQ$_{dialogic}$, and score on the AMT.

This analysis was repeated with LSHS-R$_{vis}$ as the dependent variable, and LSHS-R$_{aud}$ as an independent variable. In the first step (gender, age, LSHS-R$_{vis}$, HADS$_{anx}$, HADS$_{dep}$), the model significantly predicted LSHS-R$_{vis}$ ($F_{(5, 104)} = 11.84$, $p < .001$, $R^2 = .363$) although the only significant predictor was LSHS-R$_{aud}$ ($\beta = .497$, $p < .001$). In the second step, addition of the inner speech variables (VISQ$_{dialogic}$, VISQ$_{other}$, VISQ$_{eval}$) did not significantly improve the model ($F_{(3, 101)} = 1.44$, $p = .236$, $\Delta R^2 = .026$). In the third step, addition of the task-based measures significantly improved the model ($F_{(4, 97)} = 2.48$, $p = .049$, $\Delta R^2 = .057$), although no individual variables other than LSHS-R$_{aud}$ significantly contributed to the model. In the final model, no variables significantly predicted visual hallucination-proneness, when controlling for auditory hallucination-proneness.

3.3. Further task analysis – action monitoring

Given the unexpected findings relating to the action monitoring task, we conducted exploratory analysis to investigate the association between auditory hallucination-proneness and performance on both the self-controlled and other-controlled conditions. We hypothesized that high levels of hallucination-proneness would be associated with problems monitoring self-produced actions (i.e., a lower score on the self-controlled condition), whereas this difference would not be seen for actions controlled by someone else (i.e., no difference in score on the other-controlled condition). However, the regression analyses indicated that performance on the task (quantified as score on the self-controlled condition, minus the score on the other-controlled condition, following Williams and Happé, 2009) was *positively* associated with hallucination-proneness.

To investigate this further, we separated the sample into two groups, using a median split so that participants who scored > 9 on the LSHS-R_{aud} formed the 'high auditory hallucination-proneness' ($N = 51$) group, and those scoring ≤ 9 formed the 'low auditory hallucination-proneness' group ($N = 59$). A 2×2 mixed model ANOVA was then performed, with task condition (self-/other-controlled) and auditory hallucination-proneness (high/low) as independent variables, and number of squares correctly identified in the task as the dependent variable. There was a significant main effect of condition (self-/other-controlled) on task performance ($F_{(1, 108)} = 81.26$, $p < .001$), indicating that participants scored significantly higher on the task in the self-controlled ($M = 16.31$, $SD = 8.6$) compared to the other-controlled ($M = 9.23$, $SD = 6.4$) condition, as would be expected (see Fig. 4.2). Pairwise comparisons showed that there was an effect of auditory hallucination-proneness, with participants who scored above the median on the LSHS-R_{aud} scoring significantly

higher on the task ($M = 14.40$, $SD = 9.04$) than participants who scored below the

median ($M = 11.36$, $SD = 8.40$) ($p = .011$).
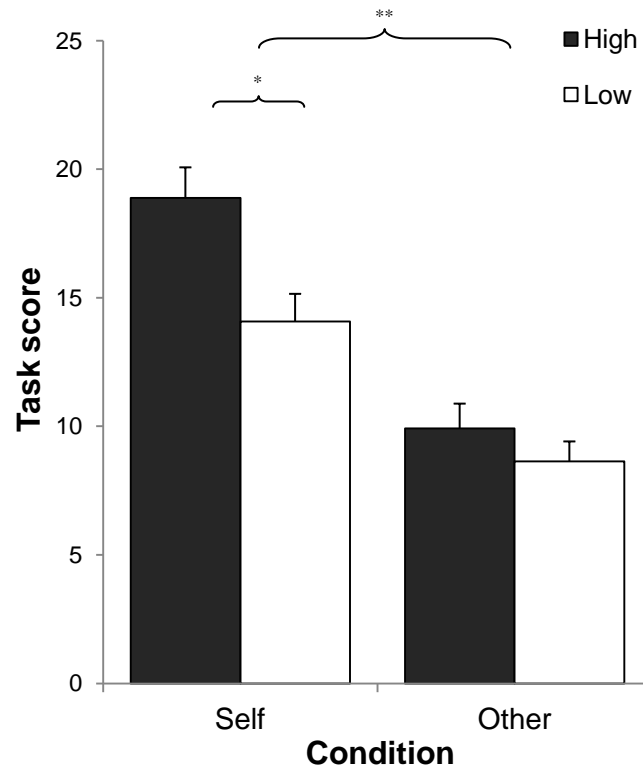
Figure 4.2: Number of squares correctly identified as controlled by the computer mouse in the action monitoring task. Split into self-controlled and other-controlled task conditions, and participants who scored above the median (high), and those who scored less than or equal to the median (low), on the LSHS-R$_{aud}$ subscale. * $p < .01$. ** $p < .001$.

There was also a significant interaction between task condition and hallucination-proneness ($F_{(1, 108)} = 4.08$, $p = .031$). Post-hoc two-tailed independent samples t-tests indicated that participants prone to auditory hallucinations scored significantly higher ($M = 18.88$, $SD = 8.48$) in the self-controlled condition than did participants who were not prone to auditory hallucinations ($M = 14.08$, $SD = 8.22$) ($t(108) = 3.01$, $p = .003$, $d = .58$). However, there was no difference between the two hallucination-

proneness groups in performance on the other-controlled condition (high: $M = 9.92$,

$SD = 6.85$; low: M = 8.63, $SD = 6.03$) ($t(108) = 1.06$, $p = .294$, $d = .20$). These

results are in the opposite direction to that predicted, and are discussed further in the

Discussion section.

4. Discussion

The present study investigated proneness to hallucinatory experiences in a non-

clinical sample of young adults, and its associations with self-reported

phenomenology of inner speech, and performance on source monitoring, reality

discrimination, and self-monitoring tasks. Results showed that, in a regression

analysis, proneness to hallucinations was independently predicted by 1) trait anxiety

(score on the 'anxiety' subscale of the HADS); 2) tendency to engage in dialogic

inner speech (score on the 'dialogic inner speech' subscale of the VISQ); 3)

performance on the auditory signal detection task; 4) number of syllables reported in

jumbled speech (the JST); and 5) performance on a task which required monitoring

self- and other-controlled hand movements (the AMT). Proneness to hallucinations

was not, however, predicted by levels of depression, other people in inner speech,

evaluative inner speech, or internal source monitoring performance.

4.1. Varieties of inner speech

Three dimensions of inner speech phenomenology, assessed by three subscales of the

VISQ, were entered into a hierarchical regression analysis: the tendency to engage in

inner speech of a dialogic nature, inner speech including the voices of other people,

and inner speech for evaluative/motivational purposes. (A fourth scale, the tendency

133

to use condensed inner speech, was not included due to the lack of an association with hallucination-proneness in previous studies.) When all other self-report and task-based measures were controlled for, only dialogic inner speech emerged as a significant predictor of hallucination-proneness (total LSHS-R score). When the auditory and visual hallucination subscales were examined separately, dialogic inner speech only predicted auditory hallucination-proneness (controlling for visual hallucinations), but not visual hallucination-proneness (controlling for auditory hallucinations). This is consistent with the findings of McCarthy-Jones and Fernyhough (2011), who found that, controlling for anxiety, depression, and proneness to visual hallucinations, only dialogic inner speech significantly predicted proneness to auditory hallucination. The findings differ, however, from the findings of Alderson-Day et al. (2014), who found that a tendency to use dialogic inner speech did not predict proneness to auditory hallucinations. This discrepancy may be due to the use of different covariates between studies. Whilst both the present study and that conducted by McCarthy-Jones and Fernyhough controlled for levels of anxiety and depression using the HADS, Alderson-Day et al. did not include these covariates, focusing instead on dissociation and self-esteem measures (the former of which was found to mediate the association between hallucination-proneness and dialogic inner speech). Alderson-Day et al. suggest that collinearity between the different inner speech measures, and an association between use of evaluative inner speech and anxiety, may mean that use of dialogic inner speech only predicts proneness to auditory hallucinations once anxiety has been partialled out.

The present findings also extend those of McCarthy-Jones and Fernyhough (2011) by showing that, when task-based measures relating to source monitoring, reality discrimination and self-monitoring were added into the model, this association was

still significant. Our findings therefore indicate that a tendency to use inner speech with the 'back-and-forth' structure of a conversation is a robust predictor of auditory hallucination-proneness in non-clinical samples, independently of the ability to distinguish between internally and externally generated perceptions, and measures of anxiety and depression. This is consistent with the observation that auditory verbal hallucinations tend to be experienced as someone else's voice, speaking to or about the individual (Nayani & David, 1996). It therefore seems that both bottom-up differences in typical inner experience (high levels of dialogic inner speech) and biases in top-down processes relating to the ability to distinguish between internally and externally generated information, contribute to the generation of auditory hallucinations. McCarthy-Jones and Fernyhough (2011) speculate that, in non-clinical samples, dialogic inner speech accompanied by source monitoring errors may explain some forms of AH. As McCarthy-Jones and Fernyhough noted, however, the positive association is not consistent with the finding (at a trend level of significance) towards a lower level of dialogic inner speech use in patients with a diagnosis of schizophrenia that experience auditory hallucinations, compared to controls (Langdon, Jones, Connaughton, & Fernyhough, 2009).

The present study further suggests that visual hallucinations seem to be associated with similar top-down biases, but not differences in the experience of inner speech (consistent with McCarthy-Jones & Fernyhough's finding in relation to inner speech). It remains to be tested whether there would be an association between proneness to visual hallucinations and phenomenology of visual mental imagery. For example, mirroring the association between dialogic inner speech and auditory hallucinations, excessively vivid visual imagery, or imagery involving the

perspective of another person, could be related to visual hallucinations. This hypothesis should be tested in the future.

4.2. Source monitoring, reality discrimination, and self-monitoring

Participants also completed tasks that have previously been used to assess internal source monitoring (the source memory task), reality discrimination (the auditory signal detection task and the jumbled speech task) and self-monitoring (the action monitoring task). The most immediately apparent finding was that, with the exception of the source memory task, these tasks appear to account for unique variation in total hallucination-proneness, and therefore do not appear to be tapping into a common underlying mechanism. Once the other variables were controlled for, three tasks were associated with proneness to hallucinations. Two of these were in the predicted direction: the auditory signal detection task, in which participants were required to generate auditory imagery of a short sentence, and simultaneously decide whether a voice was present in a burst of noise, and the jumbled speech task, in which participants were asked to respond with any words that they heard in meaningless jumbled speech.

As predicted, in the signal detection task, the results showed that a lower response bias was associated with higher LSHS-R scores; that is, hallucination-prone participants were more willing to accept that a voice was present in the noise, regardless of the actual presence/absence of speech stimuli. This is in agreement with previous studies which have associated such a bias with hallucinations, in both clinical (Bentall & Slade, 1985; Bristow, Tabraham, Smedley, Ward, & Peters, 2014) and non-clinical (Barkus, et al., 2011; Varese, Barkus, & Bentall, 2011) samples. Given our previous finding that engaging the participant in auditory

imagery during the task lowered the overall response bias, especially in those scoring

highly on the LSHS-R (see Chapter 3, this thesis) the results from this version of the

task were entered into the regression analyses, to maximize the association between

signal detection and hallucination-proneness. In Chapter 3, it was argued that this

task requires participants to distinguish between self-generated (auditory imagery)

and non-self-generated (speech stimuli) perceptions, potentially acting as an 'online'

version of a reality monitoring source memory task. If so, between-participant

variables such as the vividness of auditory imagery, or the cognitive effort associated

with generation of auditory imagery, may contribute to the response decision. For

example, a sentence that was particularly vividly imaged may be more likely to be

mistaken for an external voice, or an image that took a large amount of cognitive

effort to imagine may be less likely to externalised.

Underlying these dimensions, an efference copy system that predicts the sensory

consequences of self-generated actions may typically be (at least partially)

responsible for the experience of agency over actions (Frith, 1992; Frith, Blakemore,

& Wolpert, 2000). Imprecision in an efference copy system may be one way in

which mental imagery can be experienced as non-self-generated (Jones &

Fernyhough, 2007), and it is possible that the lack of a top-down signal to inhibit

sensory cortex may underlie, for example, excessive vividness of imagery.

Interestingly, performance on the signal detection task whilst using *auditory* imagery

did not predict scores on the auditory subscale of the LSHS-R to a greater extent

than the visual subscale. This might imply that the modality of the imagery is not

important, but rather that the task requires the participant to decide whether a

perception was self- or non-self-generated. If this were the case, a similar finding

should hold if a visual signal detection task (with concurrent visual imagery) were

used, though to our knowledge this is a hypothesis yet to be tested in relation to hallucination-proneness.

The jumbled speech task also significantly predicted total hallucination-proneness, accounting for unique variance in total LSHS-R score (but not showing specificity to the auditory or visual modalities). This task had previously been used as an ambiguous auditory stimulus in a similar manner to the signal detection tasks (Fernyhough, et al., 2007), but the observation that it accounts for unique variance in hallucination-proneness implies that, in fact, there are differing reasons for the associations. We would speculatively suggest that the jumbled speech task may serve more as a measure of the top-down effects of expectation on perception (similarly to what Bristow et al., 2014, term 'jumping to perceptions'), whereas, as discussed above, performance on signal detection tasks require the participant to distinguish between internally and externally generated perceptions. It has previously been hypothesized that suggestibility plays an important role in proneness to hallucinations (Fernyhough, et al., 2007; Young, Bentall, Slade, & Dewey, 1987), and this interpretation of our data is consistent with an important role for the effects of expectation on hallucinatory experiences. Again, the observation that this was not specific to either the auditory or visual subscales implies that this may be a mechanism common to hallucination-proneness in general.

The source memory task, aimed at assessing internal source monitoring, did not significantly predict hallucination-proneness. This finding diverges from clinical studies indicating that patients with a diagnosis of schizophrenia who hallucinate were more likely to incorrectly recall an imagined action as performed (Gawęda, et al., 2013), although it does support the claim made by Brookwell et al. (2013) that only tasks that require the participant to distinguish between self-generated and non-

self-generated items should elicit an externalising bias in hallucination-prone individuals. This argument can explain why the present study showed an association between the signal detection task (in which participants had to distinguish between self-generated auditory imagery and an external stimulus) and hallucination-proneness, but not an association between the source memory task and hallucination-proneness. It therefore seems likely that biased reality monitoring, but not internal source monitoring, may be linked to hallucination-proneness. An alternative possibility is that signal detection theory may not be the most appropriate method by which to analyse performance on memory tasks which may rely on recollective processes, such as source memory tasks (Yonelinas, 1999). For example, it is not possible, with the present data, to test the equality of variance of the signal and noise distributions, which is an important assumption of signal detection theory. This could be addressed in the future by requiring participants to provide confidence ratings during the retrieval phase of the task, which would allow this assumption to be tested.

Finally, an unexpected result of the study was that performance on the action-monitoring task, which required participants to distinguish self-generated from non-self-generated movements, was positively associated with auditory, but not visual, hallucination-proneness. Further analysis of the data showed that there was an interaction between task condition (self- or other-controlled) and auditory hallucination-proneness, such that hallucination-prone participants (scoring above the median on the LSHS-R auditory subscale) performed better than those below the median on the self-controlled condition. There was no difference between the high and low group on the other-controlled condition, implying that the observed difference between the groups was related to self-generated movement, and that the

139

highly prone participants were better able to distinguish the results of these actions from distractor movements. This was contrary to our initial hypothesis, and does not seem to support theories which suggest that hallucination-proneness is linked to aberrant monitoring of internal motor signals. This paradigm should be used in clinical samples to test whether a similar effect is found in, for example, samples of individuals with a diagnosis of schizophrenia or other psychotic disorders who do or do not hallucinate, although a relatively large literature using other experimental paradigms already indicates that hallucinations are linked to a failure to correctly monitor self-generated actions (Daprati et al., 1997; Fourneret et al., 2002; Werner, Trapp, Wüstenberg, & Voss, 2014). It is possible that, whilst tasks that require higher conscious evaluation (e.g., the signal detection task or the jumbled speech task) elicit externalising decisional biases in clinical and non-clinical hallucinators alike, deficits on tasks that draw on lower-level internal motor processes are specific to psychosis. If this were the case, an intact *self*-monitoring system may act, to some extent, as a protective factor against developing more frequent psychotic experiences; meanwhile, reality discrimination biases would be linked to a more general disposition to report hallucination-like experiences. This line of thought is admittedly speculative, and should be directly tested in the future, for example by contrasting performance on self-monitoring and reality discrimination tasks in non-clinical hallucination-prone participants and individuals with psychotic disorders who do and do not hallucinate.

4.3. Limitations and future research

Firstly, the self-report measure of hallucination-proneness (the LSHS-R) utilised in the present study, although showing satisfactory internal reliability in our sample, is not specific to any type of hallucination. Although the scale separates into auditory

and visual subscales, these have relatively few items, and within each subscale items refer to heterogeneous experiences (for example, hearing a voice or hearing music being played), which may be related to different underlying cognitive processes. Furthermore, the scale is unable to distinguish between possible subtypes of hallucinations, despite the growing consensus that phenomenologically distinct hallucinatory experiences may relate to different underlying mechanisms (Jones, 2010; McCarthy-Jones, et al., 2014). Future research should therefore aim to develop self-report measures for use with non-clinical samples that relate to different subtypes of hallucination. It might be predicted, for example, that a tendency to use dialogic inner speech would be related to auditory hallucinations that seem explicable within an inner speech framework, but not those that seem to be related to intrusions from memory (Waters et al., 2006) or hypervigilance to external stimuli (Dodgson & Gordon, 2009; Garwood, Dodgson, Bruce, & McCarthy-Jones, 2013). Development of a self-report measure that explores the prevalence of different subtypes of auditory and visual hallucinations may be more phenomenologically sensitive, and would also enable exploration of this issue, including deepening our understanding of the importance of subtypes of hallucinations.

Secondly, it is likely that other variables play moderating or mediating roles in the relationship between inner speech, source monitoring, self-monitoring and hallucination-proneness. For example, dissociation has been shown to mediate the relationship between experiencing other people as part of inner speech, and auditory hallucination-proneness (Alderson-Day et al., 2014). A key area of research relating inner speech phenomenology to hallucinations will be to investigate how and why a tendency to use certain types of inner speech are linked to higher self-reported hallucination-proneness. The dialogicality of inner speech can explain,

phenomenologically, why many AHs are experienced as the voice talking to, or conversing with, the individual. Presumably, in the case of verbal AHs, biases in reality discrimination can lead to misattribution of inner speech; however, as yet, no empirical research has investigated why a tendency to engage in dialogic inner speech would cause higher self-reported hallucination-proneness. Future research should therefore address this issue.

4.4. Conclusions

Overall, the present study has shown that a number of different variables can account for unique variance in proneness to experiences hallucinations in a non-clinical sample. We replicated previous findings that a tendency to engage in dialogic inner speech is associated with hallucination-proneness, when controlling for variables such as anxiety and depression, and extended this by showing that tasks designed to assess reality discrimination (signal detection and jumbled speech tasks) and action self-monitoring can also predict hallucinations, independently of inner speech variables. In our sample, self-monitoring for actions was positively associated with auditory hallucination-proneness (contrary to our initial predictions). Internal source monitoring, assessed using a say/imagine source memory task, was not associated with hallucination-proneness, suggesting that only self/non-self judgements may elicit the type of externalising biases associated with hallucinations.

References

Alderson-Day, B., McCarthy-Jones, S., Bedford, S., Collins, H., Dunne, H., Rooke, C., et al. (2014). Shot through with voices: dissociation mediates the relationship between varieties of inner speech and auditory hallucination proneness. *Consciousness and Cognition, 27*, 288-296.

Aleman, A., Böcker, K. B. E., Hijman, R., de Haan, E. H. F., & Kahn, R. S. (2003). Cognitive basis of hallucinations in schizophrenia: role of top-down information processing. *Schizophrenia Research, 64*(2–3), 175-185.

Barkus, E., Smallman, R., Royle, N., Barkus, C., Lewis, S., & Rushe, T. (2011). Auditory false perceptions are mediated by psychosis risk factors. *Cognitive Neuropsychiatry, 16*(4), 289-302.

Beavan, V., Read, J., & Cartwright, C. (2011). The prevalence of voice-hearers in the general population: a literature review. *Journal of Mental Health, 20*(3), 281-292.

Bentall, R. P., Baker, G. A., & Havers, S. (1991). Reality monitoring and psychotic hallucinations. *British Journal of Clinical Psychology, 30*, 213-222.

Bentall, R. P., & Slade, P. D. (1985). Reality testing and auditory hallucinations - a signal detection analysis. *British Journal of Clinical Psychology, 24*, 159-169.

Brebion, G., Ohlsen, R. I., Pilowsky, L. S., & David, A. S. (2008). Visual hallucinations in schizophrenia: confusion between imagination and perception. *Neuropsychology, 22*(3), 383-389.

Bristow, E., Tabraham, P., Smedley, N., Ward, T., & Peters, E. (2014) Jumping to perceptions and to conclusions: specificity to hallucinations and delusions. *Schizophrenia Research, 154*, 68-72.

Brookwell, M. L., Bentall, R. P., & Varese, F. (2013). Externalizing biases and hallucinations in source-monitoring, self-monitoring and signal detection studies: a meta-analytic review. *Psychological Medicine, 43*(12), 2465-2475.

Daprati, E., Franck, N., Georgieff, N., Proust, J., Pacherie, E., Dalery, J., et al. (1997). Looking for the agent: an investigation into consciousness of action and self-consciousness in schizophrenic patients. *Cognition, 65*(1), 71-86.

Ditman, T., & Kuperberg, G. R. (2005). A source-monitoring account of auditory verbal hallucinations in patients with schizophrenia. *Harvard Review of Psychiatry, 13*(5), 280-299.

Dodgson, G., & Gordon, S. (2009). Avoiding false negatives: are some auditory hallucinations an evolved design flaw? *Behavioural and Cognitive Psychotherapy, 37*(3), 325-334.

Fernyhough, C. (1996). The dialogic mind: a dialogic approach to the higher mental functions. *New Ideas in Psychology, 14*(1), 47-62.

Fernyhough, C. (2004). Alien voices and inner dialogue: towards a developmental account of auditory verbal hallucinations. *New Ideas in Psychology, 22*(1), 49-68.

Fernyhough, C., Bland, K., Meins, E., & Coltheart, M. (2007). Imaginary companions and young children's responses to ambiguous auditory stimuli: implications for typical and atypical development. *Journal of Child Psychology and Psychiatry, 48*(11), 1094-1101.

Ford, J. M., Roach, B. J., Faustman, W. O., & Mathalon, D. H. (2007). Synch before you speak: auditory hallucinations in schizophrenia. *American Journal of Psychiatry, 164*(3), 458-466.

Fourneret, P., de Vignemont, F., Franck, N., Slachevsky, A., Dubois, B., & Jeannerod, M. (2002). Perception of self-generated action in schizophrenia. *Cognitive Neuropsychiatry, 7*(2), 139-156.

Franck, N., Rouby, P., Daprati, E., Daléry, J., Marie-Cardine, M., & Georgieff, N. (2000). Confusion between silent and overt reading in schizophrenia. *Schizophrenia Research, 41*, 357-364.

Franck, N., Farrer, C., Georgieff, N., Marie-Cardine, M., Dalery, J., d'Amato, T., et al. (2001). Defective recognition of one's own actions in patients with schizophrenia. *American Journal of Psychiatry, 158*(3), 454-459.

Frith, C. D. (1992). *The Cognitive Neuropsychology of Schizophrenia*. Hove: Lawrence Erlbaum Associates.

Frith, C. D., Blakemore, S.-J., & Wolpert, D. M. (2000). Explaining the symptoms of schizophrenia: abnormalities in the awareness of action. *Brain Research Reviews, 31*(2–3), 357-363.

Frith, C. D., & Done, D. J. (1989). Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action. *Psychological Medicine, 19*(2), 359-363.

Garwood, L., Dodgson, G., Bruce, V., & McCarthy-Jones, S. (2013). A preliminary investigation into the existence of a hypervigilance subtype of auditory hallucination in people with psychosis. *Behavioural and Cognitive Psychotherapy, FirstView*, 1-11.

Gawęda, Ł., Woodward, T. S., Moritz, S., & Kokoszka, A. (2013). Impaired action self-monitoring in schizophrenia patients with auditory hallucinations. *Schizophrenia Research, 144*(1–3), 72-79.

Heinks-Maldonado, T. H., Mathalon, D. H., Houde, J. F., Gray, M., Faustman, W. O., & Ford, J. M. (2007). Relationship of imprecise corollary discharge in schizophrenia to auditory hallucinations. *Archives of General Psychiatry, 64*(3), 286-296.

Johns, L. C., & van Os, J. (2001). The continuity of psychotic experiences in the general population. *Clinical Psychology Review, 21*(8), 1125-1141.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*(1), 3-28.

Jones, S. R. (2010). Do we need multiple models of auditory verbal hallucinations? Examining the phenomenological fit of cognitive and neurological models. *Schizophrenia Bulletin, 36*(3), 566-575.

Jones, S. R., & Fernyhough, C. (2007). Thought as action: inner speech, self-monitoring, and auditory verbal hallucinations. *Consciousness and Cognition, 16*(2), 391-399.

Langdon, R., Jones, S. R., Connaughton, E., & Fernyhough, C. (2009). The phenomenology of inner speech: comparison of schizophrenia patients with auditory verbal hallucinations and healthy controls. *Psychological Medicine, 39*(4), 655-663.

Larøi, F., Van der Linden, M., & Marczewski, P. (2004). The effects of emotional salience, cognitive effort and meta-cognitive beliefs on a reality monitoring task in hallucination-prone subjects. *British Journal of Clinical Psychology, 43*, 221-233.

McCarthy-Jones, S. R., & Fernyhough, C. (2011). The varieties of inner speech: links between quality of inner speech and psychopathological variables in a sample of young adults. *Consciousness and Cognition, 20*(4), 1586-1593.

McCarthy-Jones, S. R., Thomas, N., Strauss, C., Dodgson, G., Jones, N., Woods, A., et al. (2014). Better than mermaids and stray dogs? Subtyping auditory verbal hallucinations and its implications for research and practice. *Schizophrenia Bulletin, 40*(S4), S275-S284.

Morin, A., & Michaud, J. (2007). Self-awareness and the left inferior frontal gyrus: inner speech use during self-related processing. *Brain Research Bulletin, 74*(6), 387-396.

Morrison, A. P., Wells, A., & Nothard, S. (2000). Cognitive factors in predisposition to auditory and visual hallucinations. *British Journal of Clinical Psychology, 39*(1), 67-78.

Nayani, T. H., & David, A. S. (1996). The auditory hallucination: a phenomenological survey. *Psychological Medicine, 26*(1), 177-189.

Stephane, M., Kuskowski, M., McClannahan, K., Surerus, C., & Nelson, K. (2010). Evaluation of speech misattribution bias in schizophrenia. *Psychological Medicine, 40*(5), 741-748.

Varese, F., Barkus, E., & Bentall, R. P. (2011). Dissociative and metacognitive factors in hallucination-proneness when controlling for comorbid symptoms. *Cognitive Neuropsychiatry, 16*(3), 193-217.

Varese, F., Barkus, E., & Bentall, R. P. (2012). Dissociation mediates the relationship between childhood trauma and hallucination-proneness. *Psychological Medicine, 42*(5), 1025-1036.

Vercammen, A., de Haan, E. H. F., & Aleman, A. (2008). Hearing a voice in the noise: auditory hallucinations and speech perception. *Psychological Medicine, 38*(8), 1177-1184.

Waters, F., Badcock, J., Michie, P., & Maybery, M. (2006). Auditory hallucinations in schizophrenia: intrusive thoughts and forgotten memories. *Cognitive Neuropsychiatry, 11*(1), 65-83.

Werner, J., Trapp, K., Wüstenberg, T., & Voss, M. (2014). Self-attribution bias during continuous action-effect monitoring in patients with schizophrenia. *Schizophrenia Research, 152*(1), 33-40.

Whitford, T. J., Ford, J. M., Mathalon, D. H., Kubicki, M., & Shenton, M. E. (2012). Schizophrenia, myelination, and delayed corollary discharges: a hypothesis. *Schizophrenia Bulletin, 38*(3), 486-494.

Williams, D., & Happé, F. (2009). Pre-conceptual aspects of self-awareness in autism spectrum disorder: the case of action-monitoring. *Journal of Autism and Developmental Disorders, 39*(2), 251-259.

Woodward, T. S., Menon, M., & Whitman, J. C. (2007). Source monitoring biases and auditory hallucinations. *Cognitive Neuropsychiatry, 12*(6), 477-494.

Young, H. F., Bentall, R., Slade, P. D., & Dewey, M. E. (1987). The role of brief instructions and suggestibility in the elicitation of auditory and visual hallucinations in normal and psychiatric subjects. *Journal of Nervous and Mental Disease, 175*(1), 41-48.

Chapter 5

The role of the superior temporal lobe in auditory false perceptions: a transcranial direct current stimulation study

The results of Chapters 3 and 4 are consistent with models that specify that hallucinations may occur when internal mental events, such as inner speech, are misattributed to an external source. Chapter 3 indicated that usage of auditory verbal mental imagery decreased response bias on an auditory signal detection paradigm, whilst Chapter 4 indicated that this is independent of performance on various other source monitoring and self-monitoring tasks, and self-reported phenomenology of inner speech. Drawing on previous neuroimaging evidence, Chapter 5 investigates the neural mechanisms underlying performance on auditory signal detection; specifically, using transcranial direct current stimulation to investigate the role of the left posterior superior temporal gyrus in the tendency to make false alarm responses.

Abstract

Neuroimaging has shown that a network of cortical areas, which includes the superior temporal gyrus, is active during auditory verbal hallucinations (AVHs). In the present study, healthy, non-hallucinating participants ($N = 30$) completed an auditory signal detection task, in which participants were required to detect a voice in short bursts of white noise, with the variable of interest being the rate of false auditory verbal perceptions. This paradigm was coupled with transcranial direct current stimulation, a noninvasive brain stimulation technique, to test the involvement of the left posterior superior temporal gyrus in the creation of auditory false perceptions. The results showed that increasing the levels of excitability in this region led to a higher rate of 'false alarm' responses than when levels of excitability were decreased, with false alarm responses in a sham stimulation condition lying at a mid-point between anodal and cathodal stimulation conditions. There were also corresponding changes in signal detection parameters. These results are discussed in terms of prominent cognitive neuroscientific theories of AVHs, and potential future directions for research are outlined.

1. Introduction

Auditory verbal hallucinations (AVHs) are the experience of hearing a voice in the absence of any speaker. Although experienced by between 60-80% of people with a diagnosis of schizophrenia (Sartorius et al., 1986), the experience is also reported by approximately 1.5-3% of the general population (Beavan, Read & Cartwright, 2011; Tien, 1991). Neuroimaging findings relating to AVHs have been variable, but tend to show that AVHs coincide with activation in areas of the temporal lobe such as the superior temporal gyrus (STG), and frontal lobe areas such as the inferior frontal gyrus (IFG) and anterior cingulate cortex (ACC) (Allen et al., 2012).

The STG encompasses primary auditory cortex (PAC), as well as secondary auditory cortices such as Wernicke's area/the temporoparietal junction (TPJ), and the planum temporale (PT). Due to its importance in auditory processing, the role of the STG in AVHs (and associated cognitive mechanisms), particularly in the left hemisphere, has been extensively studied. For example, repeated measurements have shown tonic hyperactivity in left STG in patients with a diagnosis of schizophrenia who experience AVHs (Homan et al., 2013). Meta-analytic findings show that, in people who experience AVHs, PAC shows reduced activation to external auditory stimuli, but increased activation to internally generated information such as AVHs (Kompus, Westerhausen & Hugdahl, 2011). Additionally, patients with a diagnosis of schizophrenia show reduced attenuation in auditory cortex when using inner speech (Simons et al., 2010), and reduced attenuation in somatosensory cortex when experiencing tactile stimulation (Shergill et al., 2014). These findings may reflect failures of internal forward models to successfully attenuate activity in response to self-produced actions (Ford & Mathalon, 2005), and/or biased attentional processes (Kompus et al., 2011). Finally, using offline repetitive transcranial magnetic

stimulation (rTMS) or transcranial direct current stimulation (tDCS) to decrease activity in Wernicke's area (left posterior STG) as a treatment protocol has been shown to reduce the frequency of AVHs (Brunelin et al., 2012; Hoffman et al., 2013; Slotema, Blom, van Lutterveld, Hoek & Sommer, 2013), possibly due to effects on activity in other auditory cortical areas in the left STG (Kindler et al., 2013).

The above evidence suggests that the left pSTG plays a crucial role in the generation and/or experience of AVHs. This is in concordance with neuroimaging evidence suggesting that, among other areas, the superior temporal gyrus is active in the neurotypical brain during verbal self-monitoring (Allen et al., 2007; McGuire et al., 1995), and when a voice is falsely detected in white noise (Barkus, Stirling, Hopkins, McKie & Lewis, 2007), an error that people who experience AVHs make more often (Brookwell, Bentall & Varese, 2013). Nevertheless, the majority of available evidence regarding the role of the STG comes from fMRI and, due to the inherently correlational nature of neuroimaging, it is hard to draw conclusions about the causality of the role of this brain area in AVHs.

Whilst attempts to treat AVHs using neurostimulation of STG or TPJ are suggestive of the critical importance of these regions, and of surrounding auditory cortical areas (Moseley, Fernyhough, & Ellison, 2013; Kindler et al., 2013), it remains to be determined how neural activations relate to underlying cognitive mechanisms. For example, if the STG is causally involved in the genesis of AVHs, it should be possible to both increase and decrease AVH frequency by modulating the level of activity accordingly. Whilst this is clearly not possible in a clinical sample due to ethical issues, one previous approach has been to use a signal detection task, in which healthy participants are asked to listen to bursts of white noise, and respond whether they believe a voice is present (Bentall & Slade, 1985). This approach

149

enables an analysis of 'correct' perceptions, as well as 'false' perceptions (or 'false alarm' responses).

Previous research suggests that individuals with a diagnosis of schizophrenia who hallucinate, and non-clinical participants who report more frequent hallucinatory experiences, are more likely to falsely perceive a voice in the noise (Barkus et al., 2011; Brookwell, Bentall & Varese, 2013; Varese, Barkus & Bentall, 2011). These studies employ signal detection analysis, and suggest that this finding is due to a difference in response bias (i.e., how willing participants are to accept that an ambiguous stimuli is present) between hallucinators and non-hallucinators, rather than a change in sensitivity to the task (the ability to distinguish between signal and noise). This is important, as it implies that individuals who experience AVHs do not have a 'deficit' on the task, but instead simply exhibit a different style of responding. However, in a study by Vercammen, de Haan & Aleman (2008) using a similar paradigm, participants who experienced AVHs showed both a lower response bias *and* lowered sensitivity to the task, suggesting that the group differences may be more complex than a response bias. Of equal importance, false perceptions on this task are associated with high levels of activation in, among other areas, the STG (Barkus et al., 2007), even compared to correct perceptions of a voice in the noise. This suggests that high levels of activity in the STG might be associated with false alarm responses in this task, perhaps reflecting a tendency to misattribute internal, self-generated processes to an external source, as in AVHs.

Nevertheless, as discussed, evidence that activity in the STG is the *cause* of false alarm responses in a signal detection task is lacking. To address this, we utilised a form of non-invasive brain stimulation, transcranial direct current stimulation (tDCS), to modulate excitability in the left posterior STG (pSTG) of non-clinical,

non-hallucinating participants. tDCS involves running a weak electrical current between two electrodes in contact with the participant's scalp, depolarizing (anodal) or hyperpolarizing (cathodal) membrane potentials of underlying neurons, resulting in a decrease in potential activity under the cathode and an increase in potential activity under the anode (Nitsche & Paulus, 2000). Furthermore, once stimulation has stopped, a reduction in GABA concentration under the anodal electrode and glutamate concentration under the cathodal electrode can be observed (Stagg & Nitsche, 2011), as well as short-lasting behavioural effects (Hummel & Cohen, 2006).

There are two main advantages of using non-clinical samples to study hallucination-like experiences: 1) results are not confounded by anti-psychotic medication or additional symptoms of psychosis; 2) it would not be ethical to attempt to increase cortical excitability in a population which may already experience potentially pathological over-activity in superior temporal regions. Our objective was to test whether modulating excitability in left pSTG would lead to a change in the number of false perceptions that participants would make on an auditory signal detection task. Specifically, given findings that levels of activity in this region are related to both AVHs and false perceptions on auditory signal detection, we hypothesized that increasing the excitability of the posterior STG using anodal stimulation would lead to an increase in false alarms, whereas decreasing excitability using cathodal stimulation would lead to a decrease in the number of false alarms.

2. Materials and Methods

2.1. Participants

The sample consisted of 30 right-handed participants (7 males, 23 females), aged 18-26 ($M = 20.6$, $SD = 2.67$). Participants were considered ineligible to take part if they reported any hearing problems, or any history of neurological or psychiatric disorder. All gave written informed consent in accordance with the Declaration of Helsinki, and ethical approval was provided by Durham University Ethics Committee. Participants were paid £15 for participation, and were naive to the aim of the study, simply being told that the study was investigating 'auditory perception'.

2.2. Signal detection task

The stimuli used in the signal detection task were similar to those used by Barkus et al. (2007; 2011), in which participants were asked to detect a voice stimulus embedded in white noise. The voice stimuli were identical to those used by Barkus et al.; a neutral, androgynous voice reading text from an instruction manual, which was segmented into 1-second clips. To set the volume levels in the task, we ran a small pilot study ($N = 8$, none of whom took part in the main study), in which participants listened to a continuous burst of white noise, within which the voice clips were played, at a gradually ascending volume level. Participants were simply asked to respond with a button press when they heard a voice, and each pilot participant's threshold was defined as the point at which they heard three consecutive voices. For the main task, we then set the volume levels at the point at which 100%, 75%, 50% and 25% of participants in the pilot study consistently detected the voices (henceforth referred to as volume levels 4, 3, 2 and 1, respectively).

The stimuli for the main task consisted of 144 5-second bursts of white noise. During 80 bursts, a voice was present for the middle 1 second ('voice-present' trials). In the voice-present trials, voices were played at one of the four volume levels, which were kept constant across all participants (a requirement of the analysis, based on signal detection theory). The remaining 64 'voice-absent' trials consisted of the white noise, with no embedded voice. Each burst was followed by 3 seconds of silence, in which the participant was instructed to respond with a button press whether they believed a voice was present in the noise (yes/no). The stimuli were pseudo-randomly ordered, so that none of the five possible trial types (voice-absent, plus four voice-present volume levels) was presented more than three times in a row. Participants were not informed how often a voice was likely to be present, but were told that voices may be present at a variety of volumes. The task was separated into two blocks, each lasting 576 seconds, with a 5 minute break between the blocks.

2.3. Transcranial direct current stimulation

Participants received 15 minutes of tDCS, using a Magstim Eldith DC stimulator. A 1.5mA current was delivered to the first 14 participants, but for the final 16 participants this was decreased to 1mA, after two participants experienced a mild headache following stimulation. The current was delivered through rubber electrodes placed in saline-soaked sponges, held in place by two rubber straps. One electrode (5 x 5cm = 25cm^2) was positioned over the left posterior superior temporal gyrus (pSTG), over electrode site CP5 according to the EEG 10-20 system. This system ensures that the electrode montage is adjusted for differing head sizes between participants, and has been used previously to target the superior temporal gyrus, and more specifically, Wernicke's area (e.g. You, Kim, Chun, Jung & Park, 2011). The second electrode (5 x 7cm = 35cm^2) was positioned above the right eye, as in other

tDCS studies (e.g. Ball, Lane, Smith & Ellison, 2013; Ellison et al., 2014). A contralateral location was chosen as this is the most commonly used in the tDCS literature (Nitsche et al., 2008). The difference in electrode size ensured that the stimulation under the superior temporal electrode reached a higher current density than under the larger electrode. There were three stimulation conditions over the pSTG: anodal, cathodal and sham stimulation. Each participant received each type of stimulation in separate sessions, with each session separated, where possible, by 7 days (mean no. days between Sessions 1–2 = 7.47, SD = 1.55, range = 6–14; mean no. days between Sessions 2–3 = 7.80, SD = 2.51, range = 3–14). The order in which participants received the three types of stimulation was counterbalanced, so that all six possible orders were represented equally in the sample. Anodal and cathodal stimulation of the pSTG consisted of 900 seconds (15 minutes) of stimulation, plus 8 seconds during which the strength of the stimulation gradually faded in, and 8 seconds during which it faded out. Sham stimulation consisted of the application of 30 seconds of stimulation, plus 8 seconds fade-in and 8 seconds fade-out; this method of sham stimulation ensured that the participant experienced the initial tingling sensation on the scalp associated with active stimulation, but did not receive sufficient stimulation to modulate neuronal excitability. This has been demonstrated to be an effective method of blinding participants to the stimulation condition (Gandiga, Hummel & Cohen, 2006).

2.4. Procedure

In each session, participants were seated in front of a laptop computer, and were provided with noise-cancelling earbuds (Creative EP-630), through which the stimuli were played. Pilot testing indicated that some participants preferred to close their eyes whilst completing the task; therefore, all participants were blindfolded to

prevent between-participant differences in visual input. Participants completed a

short practice trial before receiving tDCS, consisting of 8 bursts of white noise. The

first block of signal detection trials commenced 340 seconds after initiation of the

stimulation, thus ensuring that the task ended simultaneously with the stimulation.

Participants then sat quietly for a 300 second (5 minute) break, in which the

electrodes were removed from their scalp, to maximise the participant's comfort.

They then completed a second block of the signal detection trials. The first block of

trials is henceforth referred to as 'online' (as it was completed whilst active or sham

stimulation was applied), and the second block as 'offline' (completed after the

electrodes had been removed from the scalp).

2.5. Data analysis

Responses were categorised into four types: hits (voice-present trial, 'yes' response),

misses (voice-present trial, 'no' response), correct rejections (voice-absent trial, 'no'

response) and false alarms (voice-absent trial, 'yes' response). These responses are

expressed as a 'hit rate' (the percentage of voice-present trials on which the

participant correctly responded 'yes') and a 'false alarm rate' (the percentage of

voice-absent trials on which the participant incorrectly responded 'yes'). From these,

standard signal detection measures for sensitivity and response bias were calculated

for each block of trials completed by the participant. *d'*, a measure of sensitivity to

the stimulus, is defined as the difference between the standardised hit rate and false

alarm rate, with a higher score indicating an increased ability to distinguish signal

from noise. *β*, a measure of response bias, is calculated as outlined in Stanislaw and

Todorov (1999): $\beta = e\left\{\frac{Z(FA)^2 - Z(H)^2}{2}\right\}$, where *Z(FA)* corresponds to a standardised

false alarm rate, and *Z(H)* corresponds to a standardised hit rate. Lower *β* values

indicate a more 'liberal' response bias (i.e., participants are more likely to accept that a voice is present under ambiguous circumstances).

We used a $3 \times 2 \times 2$ mixed model design, with the three stimulation conditions (anodal/cathodal/sham) and two task blocks (online/offline) as within-subjects variables. We also included the two stimulation strengths (1.5mA/1mA) as a between-subjects variable, to test whether the alteration in current strength affected any potential main effect. We therefore conducted a mixed model ANOVA, using stimulation condition, task block and stimulation strength as independent variables, and false alarm rate as the dependent variable. This analysis was also conducted with signal detection measures as dependent variables (d', β), as well as the hit rate (in which volume level was included as a within-subjects variable). We then performed planned contrasts to investigate specifically how false alarm rate differed between the three conditions, using planned paired t-tests.

3. Results

3.1. Effects of tDCS on false alarm rate

Descriptive statistics for performance on the signal detection task are presented in Table 5.1. If assumptions of sphericity or homogeneity of variance were not met, then the Greenhouse-Geisser correction was applied. A $3 \times 2 \times 2$ (stimulation condition $\times$ task block $\times$ stimulation strength) mixed model ANOVA showed a significant main effect of stimulation condition on false alarm rate: $F_{(2, 56)} = 3.70$, $p = .031$, $\eta^2 = .117$. Planned comparisons (two-tailed paired samples t-tests) showed that, as predicted, the false alarm rate in the anodal stimulation condition was

156

significantly higher than in the cathodal stimulation condition ($t(29) = 2.52$, $p = .018$) (see Fig. 5.1). However, the difference between the anodal stimulation condition and sham condition did not reach significance ($t(29) = 1.54$, $p = .134$), and the sham condition did not differ significantly from the cathodal condition ($t(29) = 1.07$, $p = .30$). From these results, it is difficult to conclude whether the observed difference in false alarm rate between anodal and cathodal stimulation is due to an effect of one stimulation condition or the other (or both). However, given the similar difference in false alarm rate between anodal/sham and cathodal/sham, it seems probable that the observed effect was due to both an increase in false alarms in the anodal stimulation condition, and a decrease in the cathodal stimulation condition. To back this up, we conducted an exploratory within-subject polynomial contrast analysis, which indicated that there was a significant *linear* trend across the three conditions ($F_{(1, 28)} = 6.42$, $p = .017$) suggesting that the false alarm rate varied linearly with the type of stimulation applied.
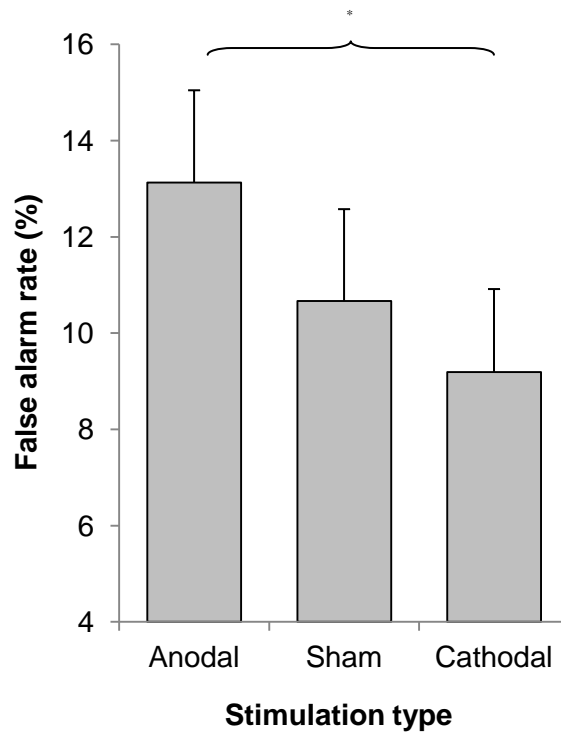
Figure 5.1: False alarm rate (%) in auditory signal detection task by stimulation condition. Error bars = 1 SEM. * $p < .05$

There was an effect of task block (online/offline) on false alarm rate: $F_{(1, 28)} = 23.68$, $p < .001$, $\eta^2 = .458$, implying that false alarm rate tended to drop between the first SDT block ($M = 13.19$, $SD = 10.33$) and the second SDT block ($M = 8.96$, $SD = 8.27$). There was no interaction between stimulation condition and task block: $F_{(1.6, 43.9)} = .396$, $p = .675$, $\eta^2 = .014$. There was also no interaction between any variables and the strength of the stimulation applied (all ps > .15), indicating that the change of current strength between participants did not change the main effect of the stimulation.

Table 5.1: Descriptive statistics for hits, false alarms, response bias and sensitivity in each stimulation condition (both task blocks) (*M, SD*). 'On' refers to performance during stimulation; 'Off' refers to performance five minutes following stimulation. FA = false alarms; β = bias; d' = sensitivity.

	Hits (%)		FA (%)		β		d'	
Type of stimulation	*On*	*Off*	*On*	*Off*	*On*	*Off*	*On*	*Off*
Anodal	59.17 (11.9)	58.08 (10.9)	15.62 (14.7)	10.64 (10.3)	3.09 (2.5)	4.22 (3.34)	1.45 (.63)	1.64 (.63)
Sham	57.92 (11.9)	56.25 (12.4)	12.05 (11.6)	9.28 (10.6)	4.32 (3.8)	4.99 (3.7)	1.6 (.61)	1.69 (.53)
Cathodal	58.08 (12.4)	58.83 (15.3)	11.69 (11.1)	6.68 (9.1)	3.61 (2.7)	5.88 (4)	1.58 (.5)	1.92 (.5)

Table 5.2: Hit rate (%) for the four different volume levels of voice embedded in the white noise. 0 = overall false alarm rate (voice-absent trials).

Volume Level	*M*	SD
4	98.6	3.67
3	66.6	17.53
2	44.5	15.88
1	22.5	12.60
0	10.6	9.86

3.2. Effects of tDCS on other signal detection measures

To explore the effect of stimulation condition on hit rate at differing voice volumes, we conducted a $3 \times 2 \times 2 \times 4$ (stimulation condition × block × stimulation strength × voice volume) mixed model ANOVA. As would be expected, there was a significant effect of volume ($F_{(2.3, 64.5)} = 361.20$, $p < .001$) on hit rate, showing that participants were more likely to correctly identify voices at higher volumes (see Table 5.2 for descriptive statistics). There was no effect of tDCS ($F_{(2, 56)} = .549$, $p = .581$) or task block ($F_{(1, 28)} = .715$, $p = .405$) on hit rate, nor any interaction between stimulation condition and task block ($F_{(2, 56)} = .581$, $p = .563$) (see Fig. 5.2). There was no interaction between voice volume and any variables (all ps > .097) or between stimulation strength and any variables (all ps > .49) From a signal detection perspective, a decrease in response bias (β) would predict an increase in false alarm rate (as observed), but also a corresponding increase in hit rate, especially in the voice stimuli presented at low volumes. Unexpectedly, our results do not support this hypothesis, since there was no effect of stimulation on overall hit rate; that is, stimulation condition affected the number of false perceptions that participants

made, but not the number of correct perceptions. There was also no interaction between voice volume and stimulation condition ($F_{(6, 174)} = .450$, $p = .844$), indicating that stimulation did not selectively affect perception of, for example, the 'below threshold' voice stimuli.
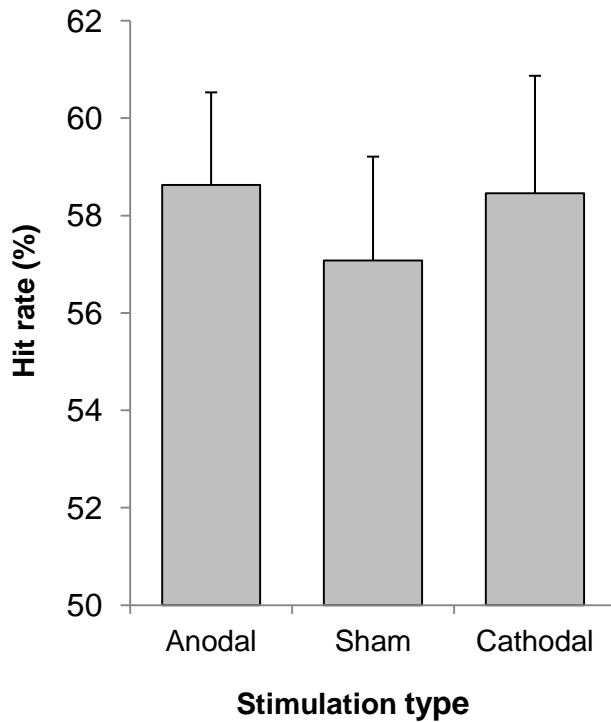
Figure 5.2: Hit rate (%) in auditory signal detection task by stimulation condition. Error bars = 1 SEM.

A $3 \times 2 \times 2$ (stimulation condition $\times$ task block $\times$ stimulation strength) mixed model ANOVA with response bias (β) as the dependent variable showed a main effect of stimulation condition, approaching significance ($F_{(2, 56)} = 2.8$, $p = .069$, $\eta^2 = .091$). Planned contrasts (two-tailed paired samples t-tests) showed that β in the anodal stimulation condition was significantly higher than in the cathodal stimulation condition ($t(29) = 2.19$, $p = .038$). (See Table 5.1 for descriptive statistics.) The

161

difference between β in the anodal and sham conditions only approached significance, (t(29 = 1.74, p = .092), but the sham condition did not differ from the cathodal condition (t(29) = .17, p = .86). There was also an effect of task block ($F_{(1, 28)}$ = 21.74, p < .001, η^2 = .437), suggesting that participants became less willing to respond that a voice was present with more experience of the task. There was no interaction between stimulation condition and task block ($F_{(2, 56)}$ = 2.39, p = .101), or any interactions between any variables and stimulation strength (all ps > .18).

A $3 \times 2 \times 2$ (stimulation condition × task block × stimulation strength) mixed model ANOVA with sensitivity (d') as the dependent variable showed that there was a main effect of tDCS on task sensitivity approaching significance: $F_{(2, 56)}$ = 3.17, p = .05. Planned contrasts (two-tailed paired samples t-tests) showed that there was a significant difference between the anodal and cathodal stimulation conditions (t(29) = 2.79, p = .010), but no significant difference between the anodal and sham conditions (t(29) = 1.05, p = .60) or the sham and cathodal conditions (t(29) = 1.25, p = .44). This difference between the anodal and cathodal conditions can be accounted for by the aforementioned larger change in false alarm rate than hit rate, as an increased false alarm rate, but stable hit rate, will lead to a decrease in sensitivity to the task. (In other words, as false alarm rate increases, the difference between the hit rate and false alarm rate decreases, leading to a lower d' score.)

4. Discussion

The present study used tDCS to study the effect of modulating cortical excitability of the left posterior superior temporal gyrus (pSTG) on rate of auditory false perceptions reported in white noise. The results showed that, as predicted, there were

significantly more false alarm responses when excitability was increased in this region using anodal stimulation, than when excitability was decreased with cathodal stimulation. The false alarm rate in the sham stimulation condition lay at a mid-point between that observed for anodal and cathodal stimulation, with the comparison with sham non-significant in each case. Signal detection analysis revealed that changes in performance due to stimulation were related to both changes in response criterion (bias) and in task sensitivity. Our findings can thus be taken to demonstrate that the left pSTG plays a role in the generation of auditory false perceptions in a non-clinical population. This is consistent with neuroimaging results showing that false alarm responses on auditory signal detection tasks are associated with over-activation of the STG, above and beyond that seen when participants correctly report hearing a voice (Barkus et al., 2007), and crucially provides evidence that this cortical area is *causally* involved in false alarm responses.

In signal detection terminology, as a result of the stimulation condition we observed a change approaching significance in response bias, where participants were more willing to respond that a voice was present when stimulation to increase the excitability of pSTG was applied. Furthermore, there was a decrease in sensitivity in the anodal stimulation condition, and an increase in sensitivity after cathodal stimulation (indicating that increasing excitability made it more difficult for participants to distinguish between the voice signal and the noise, whereas decreasing excitability made it easier). Previous studies have found that individuals who experience auditory hallucinations show a lower response bias (β), but a similar level of sensitivity (d') on auditory signal detection tasks, compared to non-hallucinating individuals (Brookwell, Bentall & Varese, 2013), which has been taken as evidence that AVHs are associated with a bias towards labelling ambiguous

percepts as external. These findings are also consistent with findings that hallucinating individuals show higher levels of activity in primary auditory cortex in response to internally generated processes such as AVHs (Kompus et al., 2011).

However, the present findings are only partially consistent with the hypothesis that differing levels of activation in the pSTG result in a differential response bias, since a change in task sensitivity was also observed as a result of stimulation condition. Nevertheless, some studies have shown a difference in sensitivity to signal detection between hallucinating and non-hallucinating groups (Vercammen et al., 2008), which may implicate broader differences in the way auditory verbal stimuli are processed in individuals prone to hallucinations. Our results suggest that higher levels of activity in the pSTG causes a bias to responding that a signal is present, but also make it more difficult to distinguish between signal and noise. Clearly, the stimulating electrodes are not simply creating an 'analogue' of individuals who experience AVHs; it is possible that, at a neural level, stimulation to increase excitability in pSTG could have reduced the signal-to-noise ratio, making it difficult to distinguish between internally and externally generated perceptions (as evidenced by the change in sensitivity).

The pSTG includes secondary auditory cortical areas such as Wernicke's area and the planum temporale (PT), as well as the temporoparietal junction (TPJ). The PT lies within secondary auditory cortex, and is preferentially active to auditory stimuli located in the external environment (Hunter et al., 2002). Thus, as well as lowering sensitivity, aberrant activation of this area may lead to a higher likelihood of a stimuli being attributed to an external source (as seen in the change in response bias). This is consistent with functional magnetic resonance imaging (fMRI) evidence showing that AVHs experienced as located in the external environment are

associated with higher levels of activation in the PT (Looijestijn et al., 2013). The TPJ, meanwhile, has been implicated in feelings of 'sensed presence', and intracranial stimulation of this area can lead to disrupted self-processing (Blanke, Ortigue, Landis & Seeck, 2002). Resting state fMRI suggests that aberrant functional connectivity between TPJ and language production areas may be associated with AVHs (Vercammen, Knegtering, den Boer, Liemburg & Aleman, 2010) which may underlie problems monitoring self-produced (inner) speech. Taken together, these findings indicate that the TPJ and other posterior temporal regions may play a key role in distinguishing between self-generated and externally-generated perceptions.

However, the area stimulated in the current study (25cm^2) does not allow us to disentangle the potential roles of specific areas of the pSTG in this task. Indeed, a limitation of the technique employed in this study is the relatively low spatial resolution, and so it is not possible to test whether differences in levels of activation of, for example, the PT or TPJ, drive the observed effect, and it is also possible that the stimulating electrode could have affected inferior parietal regions of cortex. Whilst it is not possible to resolve this issue using the current data, it should be noted that, in a neuroimaging study using a very similar task, other cortical regions such as the inferior parietal lobe were not implicated in false alarm responses (Barkus et al., 2007). In principle, stimulating electrodes with a smaller surface area could be used to investigate the role of more specific cortical regions (e.g., Borckardt et al., 2012), although these techniques are still in their infancy, and the higher current density (due to smaller electrode surface) can lead to discomfort for the participant. A combination of noninvasive neurostimulation, such as the technique used in the present study, and functional neuroimaging techniques, would allow investigation of precise cortical areas involved in false alarm responses. This would also allow

exploration of effects that may be distal to the stimulating electrode, and potentially part of a network of cortical areas involved in the genesis of auditory false perceptions.

It is, furthermore, possible that tDCS could affect functional interactions between pSTG and other regions of a cortical network thought to be involved in auditory false perceptions (for example, anterior cingulate regions involved in error detection or inferior frontal regions involved in speech production; Allen et al., 2008). Alternatively, Hoffman, Fernandez, Pittman & Hampson (2011) have argued that a corticostriatal loop involving Wernicke's area, Broca's area, and the putamen may be hyperconnected in those who experience AVHs, and that this may be affected by left TPJ stimulation. Again, integration of neurostimulation and neuroimaging paradigms will enable exploration of these issues (for a recent example, see Ellison et al., 2014).

We also observed a reduction in false alarm rate between the first and second task block, regardless of stimulation condition; that is, participants were less prone to auditory false perceptions in the offline block (after stimulation) regardless of whether they had received sham stimulation or either of the active stimulations. This is in accordance with the practice effect in auditory signal detection noted by Varese et al. (2011). There was no interaction between stimulation condition and task block, implying that the increased or decreased excitability of the pSTG was still evident in the offline block of the signal detection task. This is consistent with previous findings which have indicated that, after 15 minutes of stimulation, behavioural effects can be seen for up to an hour (Hummel & Cohen, 2006). Our results also indicated that decreasing the strength of the stimulation applied, from 1.5mA to 1mA, did not significantly affect the results (that is, there was an effect of

166

stimulation on false alarm rate regardless of stimulation strength). It might be expected that, if the stimulation strength was decreased, a smaller effect size would be observed, but it is probable that the present study was not adequately powered to pick up an interaction between stimulation condition (anodal, cathodal or sham) and stimulation strength (1.5mA/1mA). Investigating this interaction was not, however, a primary aim of the study.

One possible alternative interpretation of these findings relates to the positioning of the 'reference' electrode above the right eye. This electrode likely covered anterior areas of the right prefrontal cortex, an area which has previously been implicated in source memory retrieval (Simons, Davis, Gilbert, Frith & Burgess, 2006). It is, therefore, possible that the reported effects are due to modulation of activity in this brain region; however, neuroimaging using auditory signal detection tasks did not specifically associate this brain area with any aspects of performance, whereas superior temporal regions were specifically associated with false alarm responses (Barkus et al., 2007). We also attempted to minimise the potential effect of the frontal electrode by increasing the size of the electrode, therefore decreasing the current density and lessening the potential for neuronal modulation. Whilst it seems more parsimonious to conclude that the observed effect was due to changes in activity of the pSTG, it cannot be ruled out that modulation of areas in prefrontal cortex may be responsible for the observed effect on auditory signal detection. Indeed, many other studies utilise a contralateral frontal electrode (e.g., Ball et al., 2013; You et al., 2011), and so this is an issue which pervades much tDCS research.

Regardless of limitations relating to spatial resolution, this study has provided evidence that stimulating using this electrode montage can affect the number of auditory false perceptions on a signal detection task. This has implications for the

potential of using neurostimulation as a treatment option, with studies attempting to reduce activity in the posterior STG suggesting that this may be an efficacious treatment option to reduce their frequency (Slotema et al., 2013). It has recently been suggested that modulation of a cortical network important in self/reality monitoring and inner speech may underlie the therapeutic effect of neurostimulation (Moseley, Fernyhough, & Ellison, 2013), although evidence concerning the effect of neurostimulation on reality monitoring is limited. This would, however, be consistent with the present results, and with neuroimaging findings relating to activity in the STG during source memory and signal detection tasks (Barkus et al., 2007; Sugimori, Mitchell, Raye, Greene & Johnson, 2014), which imply that higher levels of activation may lead to perceptions being labelled as external. An interesting avenue for future research would be to investigate whether stimulation of the pSTG would affect response biases in detecting stimuli other than voices. It has been relatively well established that individuals that hallucinate are more likely to misattribute auditory verbal material (Brookwell et al., 2013), but less research has investigated other modalities. Gawęda, Woodward, Moritz, and Kokoszka (2013) found that patients with a diagnosis of schizophrenia who hallucinated showed a bias towards responding that imagined actions were actually performed, indicating that response biases may not be specific to voices, or, indeed, auditory stimuli.

Future research should therefore attempt to establish the precise relationship between reality monitoring (distinguishing between internally and externally generated perceptions) and the auditory signal detection task used here, as well as the specificity of the effect to auditory verbal material. Previous literature has tended to assume that the differential response bias shown by hallucinating participants in auditory signal detection is linked to reality monitoring mechanisms (Bentall &

Slade, 1985; Brookwell, Bentall & Varese, 2013). One could argue that a false alarm response must, by definition, be an internal mental event misattributed to an external source (i.e., the white noise), and this idea is supported by meta-analytic findings, which show a similar effect size between hallucinating and non-hallucinating samples for response biases on source memory tasks and auditory signal detection tasks (Brookwell et al., 2013). Here, we have shown that modulating excitability of left pSTG can alter the false alarm rate, and it is possible that this may be due to modulation of activity in areas important for reality monitoring, but it is not clear whether this finding would be specific to language. Future research should aim to establish whether neurostimulation of pSTG can have a similar effect on other tasks purported to test reality monitoring.

In conclusion, this study demonstrates that modulating activity in the pSTG can affect the number of false alarm responses that participants make when asked to detect speech in white noise. These results are consistent with theories that specify an important role for the pSTG in mechanisms that distinguish between internally and externally generated perceptions. It also provides a mechanism through which modulation of excitability of this cortical region may reduce frequency of AVHs, in those that seek help with anomalous experiences.

References

Allen, P., Amaro, E., Fu, C. H. Y., Williams, S. C. R., Brammer, M. J., Johns, L. C., et al. (2007). Neural correlates of the misattribution of speech in schizophrenia. *British Journal of Psychiatry, 190*, 162-169.

Allen, P., Larøi, F., McGuire, P. K., & Aleman, A. (2008). The hallucinating brain: a review of structural and functional neuroimaging studies of hallucinations. *Neuroscience & Biobehavioral Reviews, 32*(1), 175-191.

Allen, P., Modinos, G., Hubl, D., Shields, G., Cachia, A., Jardri, R., et al. (2012). Neuroimaging auditory hallucinations in schizophrenia: from neuroanatomy to neurochemistry and beyond. *Schizophrenia Bulletin, 38*(4), 695-703.

Ball, K., Lane, A., Smith, D., & Ellison, A. (2013). Site-dependent effects of tDCS uncover dissociations in the communication network underlying the processing of visual search. *Brain Stimulation, 6*, 959-965.

Barkus, E., Smallman, R., Royle, N., Barkus, C., Lewis, S., & Rushe, T. (2011). Auditory false perceptions are mediated by psychosis risk factors. *Cognitive Neuropsychiatry, 16*(4), 289-302.

Barkus, E., Stirling, J., Hopkins, R., McKie, S., & Lewis, S. (2007). Cognitive and neural processes in non-clinical auditory hallucinations. *British Journal of Psychiatry, 191*, S76-S81.

Beavan, V., Read, J., & Cartwright, C. (2011). The prevalence of voice-hearers in the general population: a literature review. *Journal of Mental Health, 20*(3), 281-292.

Bentall, R. P., & Slade, P. D. (1985). Reality testing and auditory hallucinations - a signal detection analysis. *British Journal of Clinical Psychology, 24*, 159-169.

Blanke, O., Ortigue, S., Landis, T., & Seeck, M. (2002). Neuropsychology: stimulating illusory own-body perceptions - the part of the brain that can induce out-of-body experiences has been located. *Nature, 419*(6904), 269-270.

Borckardt, J. J., Bikson, M., Frohman, H., Reeves, S. T., Datta, A., Bansal, V., et al. (2012). A pilot study of the tolerability and effects of high-definition transcranial direct current stimulation (HD-tDCS) on pain perception. *The Journal of Pain, 13*(2), 112-120.

Brookwell, M. L., Bentall, R. P., & Varese, F. (2013). Externalizing biases and hallucinations in source-monitoring, self-monitoring and signal detection studies: a meta-analytic review. *Psychological Medicine, 43*(12), 2465-2475.

Brunelin, J., Mondino, M., Gassab, L., Haesebaert, F., Gaha, L., Suaud-Chagny, M. F., et al. (2012). Examining transcranial direct-current stimulation (tDCS) as a treatment for hallucinations in schizophrenia. *American Journal of Psychiatry, 169*(7), 719-724.

Ellison, A., Ball, K. L., Moseley, P., Dowsett, J., Smith, D. T., Weis, S., et al. (2014). Functional interaction between right parietal and bilateral frontal cortices during visual search tasks revealed using functional magnetic imaging and transcranial direct current stimulation. *PLoS One, 9*(4), e93767.

Ford, J. M., & Mathalon, D. H. (2005). Corollary discharge dysfunction in schizophrenia: can it explain auditory hallucinations? *International Journal of Psychophysiology, 58*(2-3), 179-189.

Gandiga, P. C., Hummel, F. C., & Cohen, L. G. (2006). Transcranial DC stimulation (tDCS): a tool for double-blind sham-controlled clinical studies in brain stimulation. *Clinical Neurophysiology, 117*(4), 845-850.

Gawęda, Ł., Woodward, T., Moritz, S., & Kokoszka, A. (2013). Impaired action self-monitoring in schizophrenia patients with auditory hallucinations. *Schizophrenia Research, 144*, 72-79.

Hoffman, R. E., Fernandez, T., Pittman, B., & Hampson, M. (2011). Elevated functional connectivity along a corticostriatal loop and the mechanism of auditory/verbal hallucinations in patients with schizophrenia. *Biological Psychiatry, 69*(5), 407-414.

Hoffman, R. E., Wu, K., Pittman, B., Cahill, J. D., Hawkins, K. A., Fernandez, T., et al. (2013). Transcranial magnetic stimulation of Wernicke's and right homologous sites to curtail "voices": a randomized trial. *Biological Psychiatry, 73*(10), 1008-1014.

Homan, P., Kindler, J., Hauf, M., Walther, S., Hubl, D., & Dierks, T. (2013). Repeated measurements of cerebral blood flow in the left superior temporal gyrus reveal tonic hyperactivity in patients with auditory verbal hallucinations: a possible trait marker. *Frontiers in Human Neuroscience,* 7:304. doi: 10.3389/fnhum.2013.00304.

Hummel, F. C., & Cohen, L. G. (2006). Non-invasive brain stimulation: a new strategy to improve neurorehabilitation after stroke? *Lancet Neurology, 5*(8), 708-712.

Hunter, M. D., Griffiths, T. D., Farrow, T. F. D., Zheng, Y., Wilkinson, I. D., Hegde, N., et al. (2002). A neural basis for the perception of voices in external auditory space. *Brain, 126*(1), 161-169.

Jones, S. R., & Fernyhough, C. (2007). Thought as action: inner speech, self-monitoring, and auditory verbal hallucinations. *Consciousness and Cognition, 16*(2), 391-399.

Kindler, J., Homan, P., Jann, K., Federspiel, A., Flury, R., Hauf, M., et al. (2013). Reduced neuronal activity in language-related regions after transcranial magnetic stimulation therapy for auditory verbal hallucinations. *Biological Psychiatry, 73*(6), 518-524.

Kompus, K., Westerhausen, R., & Hugdahl, K. (2011). The "paradoxical" engagement of the primary auditory cortex in patients with auditory verbal hallucinations: a meta-analysis of functional neuroimaging studies. *Neuropsychologia, 49*(12), 3361-3369.

Looijestijn, J., Diederen, K. M. J., Goekoop, R., Sommer, I. E. C., Daalman, K., Kahn, R. S., et al. (2013). The auditory dorsal stream plays a crucial role in projecting hallucinated voices into external space. *Schizophrenia Research, 146*(1–3), 314-319.

McGuire, P. K., David, A. S., Murray, R. M., Frackowiak, R. S. J., Frith, C. D., Wright, I., et al. (1995). Abnormal monitoring of inner speech: a physiological basis for auditory hallucinations. *The Lancet, 346*(8975), 596-600.

Moseley, P., Fernyhough, C., & Ellison, A. (2013). Auditory verbal hallucinations as atypical inner speech monitoring, and the potential of neurostimulation as a treatment option. *Neuroscience & Biobehavioral Reviews, 37*(10), 2794-2805.

Nitsche, M. A., Cohen, L. G., Wassermann, E. M., Priori, A., Lang, N., Antal, A., et al. (2008). Transcranial direct current stimulation: state of the art 2008. *Brain Stimulation, 1*(3), 206-223.

Nitsche, M. A., & Paulus, W. (2000). Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *The Journal of Physiology, 527*(3), 633-639.

Sartorius, N., Jablensky, A., Korten, A., Ernberg, G., Anker, M., Cooper, J. E., et al. (1986). Early manifestations and first-contact incidence of schizophrenia in different cultures: a preliminary report on the initial evaluation phase of the WHO collaborative study on determinants of outcome of severe mental disorders. *Psychological Medicine, 16*(4), 909-928.

Shergill, S. S., White, T. P., Joyce, D. W., Bays, P. M., Wolpert, D. M., & Frith, C. D. (2014). Functional magnetic resonance imaging of impaired sensory prediction in schizophrenia. *JAMA Psychiatry, 71*(1), 28-35.

Simons, C. J. P., Tracy, D. K., Sanghera, K. K., O'Daly, O., Gilleen, J., Dominguez, M. G., et al. (2010). Functional magnetic resonance imaging of inner speech in schizophrenia. *Biological Psychiatry, 67*(3), 232-237.

Simons, J. S., Davis, S. W., Gilbert, S. J., Frith, C. D., & Burgess, P. W. (2006). Discriminating imagined from perceived information engages brain areas implicated in schizophrenia. *NeuroImage, 32*(2), 696-703.

Slotema, C. W., Blom, J. D., van Lutterveld, R., Hoek, H. W., & Sommer, I. E. (2013). Review of the efficacy of transcranial magnetic stimulation for auditory verbal hallucinations. *Biological Psychiatry,* doi: 10.1016/j.biopsych.2013.09.038.

Stagg, C. J., & Nitsche, M. A. (2011). Physiological basis of transcranial direct current stimulation. *Neuroscientist, 17*(1), 37-53.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137-149.

Sugimori, E., Mitchell, K. J., Raye, C. L., Greene, E. J., & Johnson, M. K. (2014). Brain mechanisms underlying reality monitoring for heard and imagined words. *Psychological Science, 25*(2), 403-413.

Tien, A. Y. (1991). Distribution of hallucinations in the population. *Social Psychiatry and Psychiatric Epidemiology, 26*(6), 287-292.

Varese, F., Barkus, E., & Bentall, R. P. (2011). Dissociative and metacognitive factors in hallucination-proneness when controlling for comorbid symptoms. *Cognitive Neuropsychiatry, 16*(3), 193-217.

Vercammen, A., de Haan, E. H. F., & Aleman, A. (2008). Hearing a voice in the noise: auditory hallucinations and speech perception. *Psychological Medicine, 38*(8), 1177-1184.

Vercammen, A., Knegtering, H., den Boer, J. A., Liemburg, E. J., & Aleman, A. (2010). Auditory hallucinations in schizophrenia are associated with reduced functional connectivity of the temporo-parietal area. *Biological Psychiatry, 67*(10), 912-918.

You, D. S., Kim, D., Chun, M. H., Jung, S. E., & Park, S. J. (2011). Cathodal transcranial direct current stimulation of the right Wernicke's area improves comprehension in subacute stroke patients. *Brain and Language, 119*(1), 1-5.

<center>**Chapter 6**</center>

<center>**Investigating the role of superior temporal and medial prefrontal cortices in reality monitoring**</center>

The findings from Chapter 5 indicated that modulating levels of activity in the left posterior superior temporal gyrus affected the number of false perceptions on an auditory signal detection task, associated with changes in both response bias and task sensitivity. Combined with the findings from Chapters 3 and 4, this provides evidence that this cortical region is involved in external misattributions typically associated with auditory verbal hallucinations. A further question, however, is whether the purported cognitive and neural mechanisms underlying external misattributions are linked to contemporary memory models relating to reality monitoring (e.g., distinguishing imagined from perceived past events).

In Chapter 4, internal source monitoring was not found to be associated with either signal detection performance, or self-reported hallucination-proneness. However, a relatively large amount of previous research indicates that performance on source monitoring tasks that require the participant to distinguish whether an event was internally generated or externally generated (reality monitoring) is associated with hallucination-proneness in both clinical and non-clinical samples (Brookwell, Bentall, & Varese, 2013), as well as implicating similar cortical regions in external misattributions. This is in contrast to the internal source monitoring task utilised in Chapter 4, which simply required participants to distinguish between words they had imagined, and words they had spoken aloud (i.e., all the words were internally generated; see Table 4.1, Chapter 4). Using transcranial direct current stimulation, the following study therefore investigated the role of both superior temporal and

<center>173</center>

medial prefrontal cortical regions in memory for the source (internal/external) of

auditory verbal stimuli.

Abstract

'Reality monitoring' refers to the ability to distinguish between internally and externally generated information. The source monitoring framework suggests that subjective attributes relating to recalled information are used to make judgements about its origin, such as perceptual vividness and cognitive effort associated with production of the information. Recent evidence has suggested that cortical regions involved in these judgements include the anterior medial prefrontal cortex (amPFC) and the left superior temporal gyrus (STG). Biased reality monitoring performance has also been linked to the tendency to experience auditory hallucinations, with the left STG in particular being implicated in the propensity to make external misattributions (i.e., to recall internally generated information as external). The present study used transcranial direct current stimulation, a noninvasive brain stimulation technique, to transiently modulate excitability in the right amPFC and left STG, in order to investigate whether these regions play crucial roles in reality monitoring for verbal information. Participants completed a source memory task in which they were asked to recall whether previously presented words had been heard or imagined. Results indicated no difference in task performance when these regions were stimulated during the encoding stage of the task compared to a sham stimulation control condition. It is concluded, therefore, that these regions play less important roles than previously theorised in encoding (and may be of greater importance in the retrieval of source information), or that more complex functional connections between different regions underlie reality monitoring abilities.

1. Introduction

The ability to recall the origin, or 'source', of information is a crucial aspect of remembering past experiences, and has been termed 'source monitoring' (Johnson, Hashtroudi, & Lindsay, 1993). Source monitoring can be separated into various categories, depending on the sources of information that must be distinguished between. For example, internal source monitoring requires the participant to distinguish between two or more internal sources (e.g., imagining a word, or speaking a word aloud), whilst external source monitoring requires the participant to distinguish between two or more sources external to the self (e.g., whether a word was spoken by one person or another person; see Chapter 4, Table 4.1). The ability to recall whether information was externally generated (i.e., emanated from the surrounding environment) or internally generated (an action performed or imagined by oneself) has been termed 'reality monitoring' (Johnson, et al., 1993; Johnson & Raye, 1981; Mitchell & Johnson, 2009). Research falling under the source monitoring framework has investigated the specific qualitative attributes of memories that may contribute to judgements of source; for example, an event represented in memory as especially vivid (i.e., high in perceptual detail) may be more likely to be recalled as originating from the external environment (Johnson, Foley, & Leach, 1988). Conversely, a remembered event associated with high cognitive effort may be recalled as self-generated, or 'internal' (Finke, Johnson, & Shyi, 1988). Typically, reality monitoring is assessed using a source memory paradigm, in which the participant must recall whether a stimulus (e.g., a word or image) was previously presented to them, or whether they imagined/spoke the stimulus themselves (Brookwell, et al., 2013; Johnson, et al., 1993).

Cognitive neuroscientific research into reality monitoring has implicated anterior medial prefrontal cortical (amPFC) regions in distinguishing between internally and externally generated perceptions (Mitchell & Johnson, 2009; Vinogradov et al., 2006). For example, Simons, Davis, Gilbert, Frith, and Burgess (2006) showed that, during the retrieval phase of a source memory paradigm, reduced activation in the amPFC was associated with the likelihood that imagined events would be recollected as perceived. Furthermore, Buda, Fornito, Bergström, and Simons (2011) demonstrated that absence of the paracingulate sulcus (located adjacent to, and associated with differential grey matter volumes in, the medial PFC) was associated with impaired reality monitoring performance. It has been theorised that activity in PFC regions may reflect engagement of higher level cognitive operations and so should be important in the feeling of cognitive effort associated with self-generated events (Mitchell & Johnson, 2009; Sugimori, Mitchell, Raye, Greene, & Johnson, 2014). For example, drawing on evidence relating anterior medial PFC activity to, among other things, metacognition, theory of mind, and the default mode network, the 'gateway hypothesis' specifies that anterior PFC plays a role in switching between stimulus-oriented (i.e., external) and stimulus-independent (i.e., internal) thought (Burgess, Simons, Dumontheil, & Gilbert, 2005). It is therefore no surprise that amPFC is important in the ability to retrospectively distinguish between events that were external and events that were internal.

Similarly, Sugimori et al. (2014), in a paradigm that required participants to recall whether words had previously been heard or only imagined, showed that frontal regions, such as the middle frontal gyrus, were more active during encoding of words that were correctly recalled as 'imagined'. This could be interpreted as evidence that engagement of cognitive operations (reflected in MFG activity) later

acted as a cue that stimuli were self-generated. Furthermore, Sugimori et al. showed that words later recalled as 'heard' were associated with activity in speech production areas such as the inferior frontal gyrus (IFG), regardless of the original source of the word. The authors argue that the involvement of IFG in reality monitoring provides evidence that semantic and perceptual detail from speech production areas can inform the judgement of source at retrieval. Interestingly, Sugimori et al. also investigated the link between reality monitoring, cortical activation, and the tendency of the participants to experience auditory hallucinations (assessed using a self-report measure). It was found that activity in the superior temporal gyrus (STG; encompassing primary and secondary auditory cortex) when participants incorrectly recalled a word they had imagined as 'heard', was significantly correlated with the tendency to experience auditory hallucinations.

The most prominent cognitive models which attempt to explain auditory verbal hallucinations (AVHs) have linked deficits or biases in self-monitoring or reality monitoring to the tendency to report hallucinatory experiences (Bentall, Baker, & Havers, 1991; Seal, Aleman, & McGuire, 2004). These models are evidenced by studies indicating that individuals with a diagnosis of schizophrenia who hallucinate are more likely to make external misattributions on source memory tasks (that is, incorrectly recall that a self-generated word was spoken by someone else), than those that do not hallucinate (Stephane, Kuskowski, McClannahan, Surerus, & Nelson, 2010; Woodward, Menon, & Whitman, 2007). One possibility highlighted by the source monitoring framework is that individuals who hallucinate generate excessively vivid mental imagery (Aleman, Böcker, Hijman, de Haan, & Kahn, 2003), which is therefore more likely to become misattributed. An alternative, though not exclusive, possibility, is that atypical efference copy mechanisms relating

to sensory predictions of self-generated motor acts cause internally generated events to be misattributed to an external source (Ford & Mathalon, 2005; Frith, 1992). In this way, it is hypothesized that inner speech may become misattributed to an external source, and experienced as an AVH (Jones & Fernyhough, 2007).

Symptom-capture neuroimaging studies have shown that, during the experience of an AVH, areas including the left IFG, left STG and anterior cingulate cortex (ACC, which is immediately adjacent to, and often referred to as part of, the medial prefrontal cortex) are often active (Allen, Larøi, McGuire, & Aleman, 2008; Jardri, Pouchet, Pins, & Thomas, 2011). Neuroimaging has also indicated that atypical connectivity between frontal and temporal regions is associated with AVHs (Lawrie et al., 2002; Mechelli et al., 2007). The fact that AVHs are associated with similar areas to those involved in reality monitoring tasks has been taken to support theories specifying that they result from biased monitoring processes, potentially specific to inner speech, supported by evidence of activation in the left IFG during AVHs (Jardri et al., 2011). High levels of activity in the STG could reflect either high levels of vividness of inner speech (Zvyagintsev et al., 2013), and/or a failure of efference copy processes to attenuate sensory cortices in response to a self-generated action, such as inner speech (Simons et al., 2010).

Recent research has studied the effects of reducing cortical excitability in superior temporal regions as a treatment option for AVHs, with results suggesting that decreasing excitability using techniques such as offline repetitive transcranial magnetic stimulation (rTMS) or cathodal transcranial direct current stimulation (tDCS) can lead to a reduction in the frequency of AVHs (Hoffman et al., 2013; Slotema, Blom, van Lutterveld, Hoek, & Sommer, 2013). It has been suggested that, at a cognitive level, this reduction in AVH frequency may be linked to improved

self-/reality monitoring ability (Moseley, Fernyhough, & Ellison, 2013; Chapter 1).

Our previous findings suggest that modulating excitability in the superior temporal

lobe using tDCS affects performance on a signal detection task (Moseley,

Fernyhough, & Ellison, 2014; Chapter 5), a task which is thought to index a similar

mechanism to that which distinguishes between internal and external perceptions

(Brookwell et al., 2013). Furthermore, one study showed that there was a concurrent

improvement in performance on a source memory task alongside a reduction in AVH

frequency following treatment using offline 1 Hz rTMS (Brunelin et al., 2006).

These studies therefore inform two streams of research, relating to: 1) the importance

of prefrontal areas, and primary and secondary auditory cortical regions in the STG,

for reality monitoring of auditory stimuli; 2) the importance of reality monitoring

(internal/external) processes in AVH. However, as yet, no study has tested the

involvement of these brain regions in reality monitoring using neurostimulation, in a

non-clinical population.

Therefore, the present study tested the involvement of two cortical regions in reality

monitoring for auditory verbal stimuli: the right anterior medial prefrontal cortex and

the left superior temporal gyrus. We used a noninvasive brain stimulation technique,

tDCS, to increase or decrease cortical excitability whilst participants completed the

encoding stage of a source memory task. The rationale for applying stimulation

during the encoding stage was based on models of reality monitoring suggesting that

vividness of self-generated imagery, and cognitive effort associated with the

imagery, may underlie later source judgements. tDCS involves running a weak

electrical current between two electrodes placed on the participant's scalp,

depolarising or hyperpolarising the membrane potentials of underlying neurons,

under the anodal and cathodal electrode respectively. Cortical excitability is

therefore increased underneath the anode, and decreased underneath the cathode (Nitsche & Paulus, 2000). The source memory task required participants to recall whether they had heard a word, imagined hearing it, or whether the word was completely new. Participants were therefore assessed for old/new recognition, as well as source discrimination. They were also required to provide ratings for the vividness with which they had heard or imagined each word, since, as has been outlined, this is thought to be a key variable that may affect reality monitoring performance.

In this study, tDCS was applied during the encoding phase of the task, whilst participants were presented with a series of words on the screen. For each word, the participant was either asked to listen to a male voice speak the word, or to imagine hearing the same male voice speak the word. At a later stage, the participant was then required to recall, for each word, whether they believed it had been heard or imagined earlier in the task, or if they believed the word was new (had not been presented earlier). The hypothesis was that, by increasing cortical excitability in the left STG, the rate at which participants would respond that they had heard the previously presented words would increase (as in individuals who experience AVHs). We also predicted that decreasing excitability in the right amPFC would lead to a decrease in correct source judgements overall, but that old/new discrimination would not be affected.

Following Sugimori et al. (2014), participants also completed self-report measures, to test whether proneness to hallucinations was associated with source memory in each stimulation condition. Based on the results of Sugimori et al., who showed that activation in the left STG for imagined items subsequently recalled as heard correlated with hallucination-proneness, we hypothesized that external misattribution

errors would correlate with hallucination-proneness, and that this may be most evident in the temporal stimulation condition. Finally, as outlined, previous neuroimaging findings indicate a role for the amPFC in meta-cognition (Burgess et al., 2005), especially with respect to memory (Baird, Smallwood, Gorgolewski & Margulies, 2013). Participants were therefore asked to provide a meta-cognitive estimate of reality monitoring task performance, with the hypothesis that meta-cognitive estimates would be worse following right amPFC stimulation, compared to sham stimulation.

2. Materials and Methods

2.1. Participants

The sample consisted of 36 participants (9 males, 27 females), aged 18–28 ($M = 20.14$, $SD = 2.5$). All participants were right-handed, and were also considered ineligible to take part if they reported any hearing problems, or any history of neurological or psychiatric disorder (see Appendix 1 for specific eligibility criteria). Ethical approval was provided by Durham University Ethics Committee, and written informed consent was given by participants, in accordance with the Declaration of Helsinki. All participants were rewarded with a £25 gift voucher for participating, and course credits if required. The study was advertised as testing 'auditory memory'.

2.2. Source memory task

The source memory task was based upon that used by Sugimori et al. (2014). The task consisted of two stages: the encoding stage and the recall stage. The stimuli

consisted of 450 words taken from the MRC Psycholinguistic Database (Wilson, 1988), each consisting of 1-3 syllables and 4-6 letters. The words were separated into nine lists, each consisting of 50 words, matched on mean number of letters and syllables, as well as Kucera-Francis frequency, familiarity, concreteness, imageability and meaningfulness.

In the encoding stage, participants were presented with a series of 100 words, taken from two lists. One list was assigned to be 'heard', and the other list 'imagined'. Participants were presented with each word consecutively. The heard and imagined words were presented alternately, although the order in which the items from each list were presented was randomised for each participant. Immediately before each word, participants were cued with the word 'HEAR' or 'IMAGINE' in the centre of the screen. Each word was then presented for 2500ms in the centre of the screen (font: Arial, font size: 24 pt; see Fig. 6.1 for example). If the word followed the cue to hear the stimuli, it was accompanied by an auditory stimulus, of a male voice speaking the word once. If they had been asked to imagine the presented word, no auditory stimulus was presented, and the participant was required to imagine hearing the word being read out in the same male voice they had heard. Following the presentation of each word, the participant was cued to provide a rating for how vividly they had heard/imagined the word (1 = low; 2 = average; 3 = high). Specifically, participants were informed that they should rate each word for how 'clear, detailed and realistic' it had sounded. The rating screen was presented for 2.5 secs, regardless of whether a response was entered. The encoding stage therefore lasted a total of 900 seconds.

Following the encoding stage, participants were given a short break from the task (during which time the tDCS electrodes were removed from their scalp). Participants

183

were either given a break of 5 or 15 minutes. We included this between-subject variable because neurostimulation techniques such as tDCS can have after-effects beyond the period of stimulation, both at a neural (Stagg & Nitsche, 2011) and behavioural (Hummel & Cohen, 2006) level, which may have directly affected task performance in the recall phase (as opposed to indirectly through the effect on encoding that we wished to study). If any observed effect was due to after-effects of the tDCS, it would be expected to be weaker after a longer time period, as the effects of stimulation began to wear off. During the task break, participants were asked to sit quietly in the darkened room; it was not appropriate to give participants a distractor task, since this may have interacted with the effects of the tDCS.

After the task break, participants completed the recall stage of the task. In this stage, they were presented with the two lists of words included in the encoding stage, as well as a third list which had not been previously presented. In this stage, they were therefore presented with 50 words which had previously been heard, 50 words which had previously been imagined, and 50 words which had not previously been presented. The words were presented in a random order, and appeared on the screen until a response was entered. For each word, participants were asked to respond, with a button press, whether they believed the word had been heard or imagined in the first stage, or if the word was completely new.
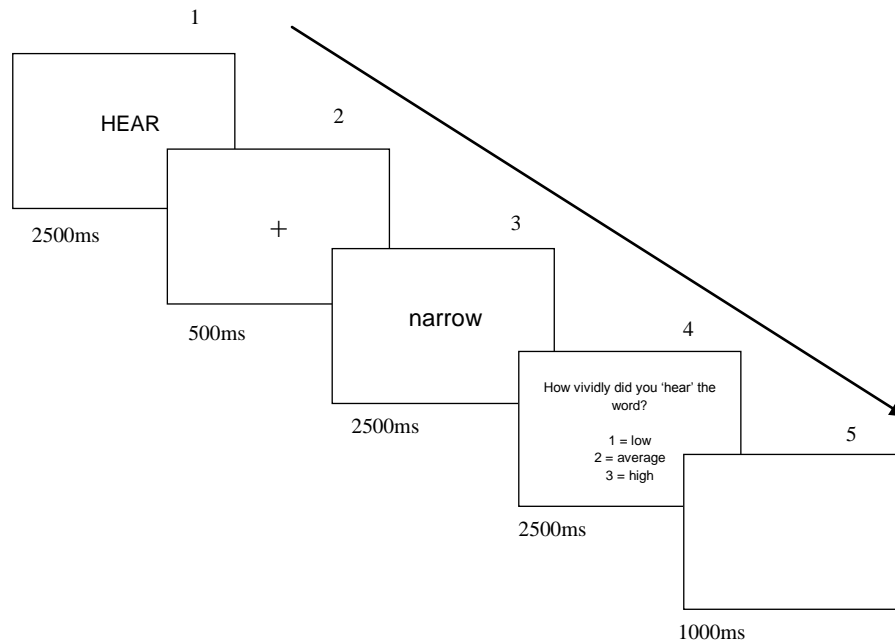
Figure 6.1: Schematic of a single trial during the encoding stage of the reality monitoring task. Participants were presented with either the word 'HEAR' or 'IMAGINE' (Screen 1), followed by a target word (Screen 3), followed by a screen which asks them to rate the vividness with which they heard or imagined the word (Screen 4).

2.3. Other measures

As well as completing the source memory task, participants were asked to complete a number of self-report measures relating to both their experience of the task, and proneness to hallucinatory experiences.

After each session, participants completed a short questionnaire which asked them the following three questions relating to the task they had just completed, requiring a response between 0–100:

1) 'Please enter an overall vividness rating, corresponding to how vividly, when asked, you think you were able to imagine each of the words';

185

2) 'Please enter an overall rating, corresponding to how often you think the voice you imagined was in the other person's voice (as opposed to your own)';

3) 'What percentage of the words do you think you correctly classified as heard/imagined/new? (chance performance = 33%)'.

The aim of the first question was to provide a retrospective estimate of the vividness with which the words were imagined, providing participants with a wider range of possible answers (0–100, as opposed to 1–3 during the task). The second question was included to provide a measure of the extent to which participants engaged in imagery of the male voice (as opposed to their own voice), as they were instructed. The aim of the third question was to assess the participant's ability to meta-cognitively estimate how well they had performed on the task.

Finally, participants completed the Revised Launay-Slade Hallucinations Scale (LSHS-R; McCarthy-Jones & Fernyhough, 2011), a self-report measure of hallucination-proneness that asks participants to rate the frequency of hallucinatory experiences (e.g., 'I have had the experience of hearing a person's voice and then found that no-one was there.'). The scale consists of 9 items, with each item scored between 1-4, summing to a possible total of 36.

2.4. Transcranial direct current stimulation

Participants received 900s (plus 8s of fade-in and 8s of fade-out) of tDCS whilst completing the encoding stage of the source memory task (see Section 2.2), using a Magstim Eldith DC Stimulator. A 1 mA current was delivered through two 5 x 5cm (25cm^2) electrodes, placed in sponges soaked in saline solution, and held in place on participants' scalp by two rubber straps. There were three separate stimulation

conditions, which all participants completed. In these, three regions were stimulated: right anterior medial PFC (amPFC), left STG, and visual area V5/MT (the latter was chosen as a control site, because it was not expected to play a role in the source memory task used in the present study). These sites were localised using the EEG 10-20 system, which adjusts for individual head size, and has previously been used to target the left STG under electrode site CP5 (Moseley et al., 2014; You, Kim, Chun, Jung, & Park, 2011) and right amPFC under electrode FP2 (Karim et al., 2010). The V5 electrode was positioned 3cm above the inion, and 6cm left of the midline, as in previous studies which have stimulated this cortical region (Antal et al., 2004).

Participants therefore received stimulation on three occasions. In one session, the cathodal electrode was positioned over the right amPFC, and the anodal electrode over the left STG (subsequently referred to as the 'temporal' condition, see Fig. 6.2a). This condition was included to investigate the effect on performance of stimulating the left STG, over and above (or in interaction with) the effect of stimulating the right amPFC. In another session, the cathodal electrode was positioned over the right amPFC, and the anodal electrode over left V5 (subsequently referred to as the 'occipital' condition, see Fig 6.2b). This was included to investigate the effect of right amPFC on task performance, without stimulation of the left STG. Finally, as a baseline condition, in another session, the electrodes were positioned as in the temporal condition, but stimulation was only applied for 30 seconds (plus 8s fade-in and 8s fade-out; subsequently referred to as the 'sham' condition). Sham stimulation is not sufficient to modulate neuronal excitability, and has previously been demonstrated to be an effective method of blinding participants to the condition (Gandiga, Hummel, & Cohen, 2006). Where

possible, sessions were separated by 7 days (mean no. days between Sessions 1-2 =

7.17, SD = 0.51, range: 7-9; mean no. days between Sessions 2-3 = 6.92, SD = 0.55,

range: 5-8). The order in which participants completed each condition was

counterbalanced.

(a)   (b)
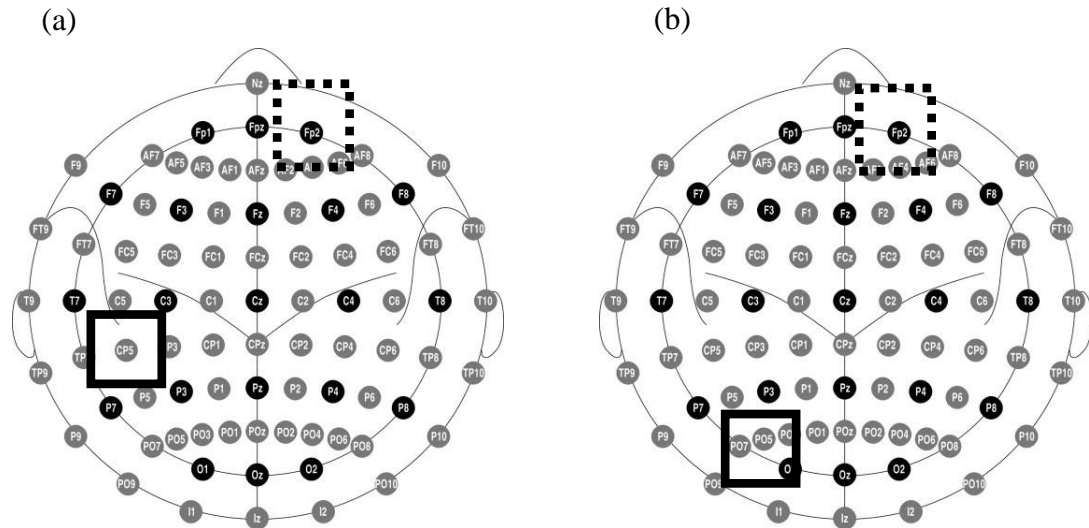
Figure 6.2: Top-down representation of electrode positioning in (a) the temporal condition and (b) the occipital condition. Dashed line = cathodal electrode; solid line = anodal electrode. (Note that the anodal electrode in the occipital condition was not positioned according to the EEG 10-10 system – see text for details.) Effects specific to the temporal condition would indicate specificity to the electrode placed over the left STG, whereas effects of both active conditions, compared to sham stimulation, would indicate an effect of the electrode placed over right amPFC. Image retrieved from https://www.kickstarter.com/projects/openbci/openbci-an-open-source-brain-computer-interface-fo/posts/712302.

2.5. Procedure

In each session, participants were seated in front of a computer and provided with

earbuds (Creative EP-630), through which the stimuli were played. Before starting

the reality monitoring task, participants listened to a short sound clip, consisting of a

male voice reading a brief passage of text (60s). This was so that the participants had

prior experience of the voice they would be asked to imagine hearing as part of the reality monitoring task. Participants then completed a short practice of both stages of the task (consisting of 6 words in the encoding stage, then 9 words in the recall stage). Participants were allowed to repeat the practice stage if they wished. The tDCS was then started, and after the 8s fade-in period, the encoding stage of the task began. After 900s, the encoding stage of the task ended, at the same time as the fade-out period (8s) of the tDCS began. The participant was asked to sit quietly for either 5 or 15 minutes, in which time the electrodes were removed from their scalp. They then completed the recall stage of the task, followed by the short questionnaire outlined in Section 2.3 (all sessions), and the hallucination-proneness measure (LSHS-R, final session only).

This procedure was kept identical across all three sessions (i.e., participants listened to the sound clip and completed the practice task each time they participated). The only difference between each session was that 1) different word lists were used in each session; 2) the stimulation condition was varied.

2.6. Data analysis

For old items (those presented in the encoding stage), within-subject variables that were entered into the analysis were stimulation condition (temporal/occipital/sham condition), presentation type (hear/imagine; representing the two types of trial that were part of the encoding stage), and response (heard/imagined; representing the source decision made by the participant in the recall stage). Interval length (5min/15min, representing the length in time between the encoding and recall stages of the task) was also included as a between-subject variable. To analyse task performance, a $3 \times 2 \times 2 \times 2$ mixed model ANOVA was therefore carried out with

number of responses as the dependent variable. The crucial effect of interest was stimulation condition × response type (with the prediction that heard responses would be higher in the temporal stimulation condition, but imagined responses would not). Power analysis indicated that, for a medium effect size ($f = .25$) and 80% power, a 3×2 (in this case, stimulation condition × response) interaction with two within-subject variables required a minimum of 30 participants.

Mean vividness ratings were calculated for each presentation type and response type, by dividing the sum of the responses after each word (1-3) by the number of each type of response. To analyse the vividness ratings, stimulation condition (temporal/occipital/sham condition), presentation type (hear/imagine) and response (heard/imagined) were included in the design as within-subject variables, as well as interval length (5/15) as a between-subjects variable. A $3 \times 2 \times 2 \times 2$ (stimulation condition × presentation type × response type × interval length) mixed model ANOVA was therefore carried out, with mean vividness rating as the dependent variable. This was to investigate whether stimulation affected the vividness with which participants reported hearing/imagining the presented words, as well as whether vividness ratings differed between the presentation types and response types. For all analyses carried out on vividness ratings, participants who made 0 of any response type ($N = 4$) were not included, because mean vividness ratings could not be calculated for all types of response.

Old/new recognition was also investigated, by entering the number of old/new errors into a $3 \times 2 \times 2$ mixed model ANOVA (stimulation condition × error direction × interval length). 'Error direction' refers to incorrect responses, either classifying an old item (a hear/imagine item in the encoding stage) as new, or a new item (not previously presented) as old. The mean vividness ratings of old words was also

analysed, to investigate whether old words correctly recalled as such (heard/imagined) were originally rated as more vivid than old words that were subsequently labelled new.

3. Results

3.1. Reality monitoring performance

Descriptive statistics for reality monitoring performance are presented in Table 6.1 and Fig. 6.3. There was no main effect of stimulation condition on number of heard and imagined responses ($F_{(2, 68)} = 0.78$, $p = .46$), showing that stimulation did not affect the tendency to label a word as heard/imagined. Stimulation condition did not interact with presentation type ($F_{(2, 68)} = 1.45$, $p = .24$), indicating that stimulation did not specifically affect reality monitoring performance on items presented as 'hear' or 'imagine'. Importantly, stimulation condition did not interact with response type ($F_{(2, 68)} = .004$, $p = .99$), indicating that altering the type of stimulation had no effect upon the tendency to recall words as heard or imagined (irrespective of the original presentation mode). Finally, stimulation condition did not interact with any other variables, or combination of variables (all ps > .16).

There was a significant main effect of presentation type ($F_{(1, 34)} = 5.46$, $p = .025$) on number of responses, indicating that items originally presented as 'hear' ($M = 17.13$, $SD = 3.3$) were more likely to be recalled as heard/imagined than 'imagine' items ($M = 16.36$, $SD = 2.9$). There was also a main effect of response type ($F_{(1, 34)} = 5.49$, $p = .025$), indicating that, other things being equal, participants were more likely to respond that they had imagined a word ($M = 17.84$, $SD = 4.2$) than that they had

191

heard a word ($M = 15.66$, $SD = 4.1$). Importantly, there was also a significant interaction between presentation type and response type ($F_{(1, 34)} = 49.51$, $p < .001$), explained by the result that participants were more likely to respond 'heard' ($M = 21.13$, $SD = 7.26$) than 'imagined' ($M = 13.14$, $SD = 4.76$) for items originally presented to 'hear' ($t(35) = 4.60$, $p < .001$), and were more likely to respond imagined ($M = 22.54$, $SD = 7.15$) than heard ($M = 10.19$, $SD = 4.68$) for items originally presented to 'imagine' ($t(35) = 7.02$, $p < .001$). (This interaction indicates that participants were able to recall the source of items at a level that was well above chance.) Apart from response type, presentation type did not interact with any variables or combination of variables (all ps $> .24$). There was a significant interaction between response type and interval length ($F_{(1, 34)} = 4.14$, $p = .049$). Further exploration using paired t-tests showed that, after a 5 minute interval, participants were more likely to respond that they had imagined a word ($M = 18.85$, $SD = 4.54$) than heard a word ($M = 14.78$, $SD = 4.70$), irrespective of the original presentation type ($t(17) = 2.53$, $p = .022$). However, after a 15 minute interval, there was no difference between the number of heard ($M = 16.54$, $SD = 3.33$) and imagined ($M = 16.82$, $SD = 3.59$) responses (irrespective of the actual source).

Table 6.1: Task performance on hear/imagine judgements on reality monitoring task (%), by interval length, response type and stimulation condition (*M, SD*)

Interval length	Presentation type	Response type	Stimulation condition		
			Temporal	Occipital	Sham
5 mins	Hear	Heard	43.00 (16.7)	40.33 (18.3)	44.11 (20.5)
		Imagined	28.22 (12.7)	26.56 (8.6)	25.89 (9.9)
	Imagine	Heard	15.89 (9.5)	16.78 (10.9)	17.22 (9.6)
		Imagined	47.33 (16.5)	46.78 (16.4)	51.44 (20.7)
15 mins	Hear	Heard	40.67 (10.4)	43.22 (14.7)	42.22 (13.3)
		Imagined	25.22 (12.0)	26.33 (13.2)	25.44 (8.9)
	Imagine	Heard	23.89 (9.1)	24.67 (9.7)	23.78 (11.2)
		Imagined	40.00 (12.2)	42.56 (11.1)	42.33 (11.8)
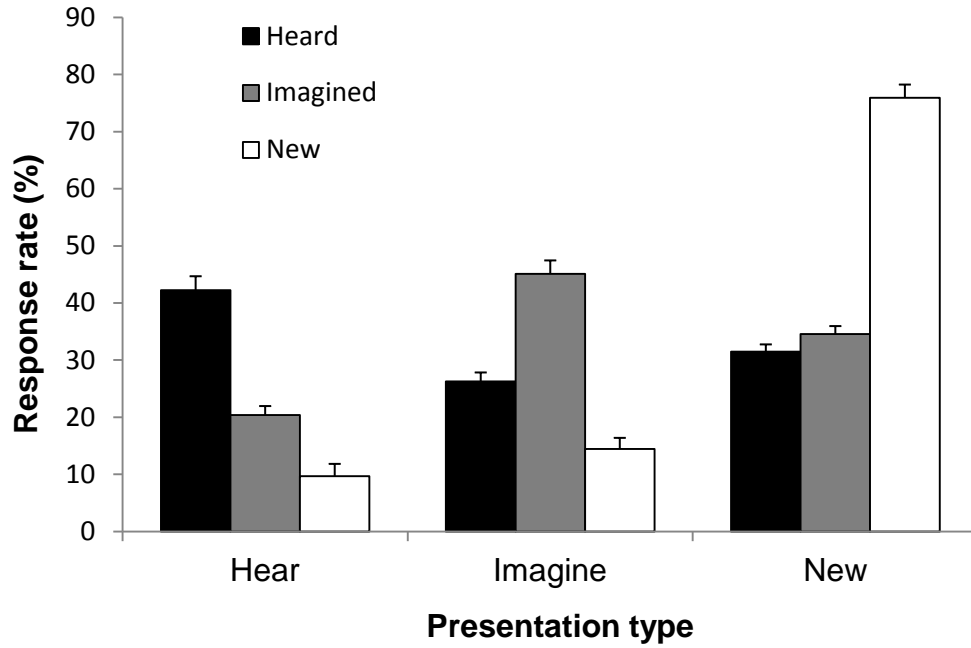
Figure 6.3: Task performance (collapsed across stimulation conditions) on the reality monitoring task. 'Presentation type' refers to whether the word was originally heard or imagined by the participant; within this, each bar represents the rate at which participants labelled words heard, imagined or new.

3.2. Effects of imagery vividness on reality monitoring performance

Descriptive statistics for vividness ratings for heard and imagined words are presented in Table 6.2. There was no main effect of stimulation condition on vividness ratings of the heard or imagined words ($F_{(2, 60)} = 1.62$, $p = .207$), nor any interaction between stimulation condition and presentation type ($F_{(2, 60)} = 0.31$, $p = .737$) or between stimulation condition and response type ($F_{(2, 60)} = 0.38$, $p = .685$).

There was a main effect of presentation type on vividness ratings ($F_{(1, 30)} = 159.02$, $p < .001$), indicating that words that were presented as 'hear' ($M = 2.76$, $SD = 0.21$) were rated as more vividly perceived than words that were presented as 'imagine' ($M = 2.00$, $SD = 0.28$), as would be expected. There was also a main effect of response

194

type ($F_{(1, 30)} = 17.77$, $p < .001$), indicating that words recalled as 'heard' ($M = 2.42$, $SD = 0.18$), irrespective of their original source, were rated as more vivid than those recalled as 'imagined' ($M = 2.34$, SD = 0.20). There was no presentation type × response type interaction ($F_{(1, 30)} = 2.24$, $p = .145$), indicating that the effect of response type on vividness rating was not different for correct or incorrect source judgements (that is, items correctly classified as heard were previously rated as higher in vividness than those incorrectly classified as imagined; and items incorrectly classified as heard were previously rated as higher in vividness than items correctly classified as imagined). Finally, there was no interaction between any other variables included in the model (all ps > .095).

Table 6.2: Mean vividness ratings for heard and imagined words, by interval length, response type and stimulation condition (*M, SD*)

Interval length	Presentation type	Response type	Stimulation condition		
			Temporal	Occipital	Sham
5 mins	Hear	Heard	2.70 (0.3)	2.77 (0.2)	2.72 (0.3)
		Imagined	2.59 (0.5)	2.76 (0.2)	2.59 (0.4)
	Imagine	Heard	1.96 (0.4)	2.07 (0.3)	2.11 (0.5)
		Imagined	1.92 (0.4)	1.96 (0.3)	1.88 (0.4)
15 mins	Hear	Heard	2.81 (0.2)	2.84 (0.1)	2.86 (0.1)
		Imagined	2.79 (0.2)	2.84 (0.2)	2.85 (0.2)
	Imagine	Heard	2.05 (0.3)	2.06 (0.5)	2.06 (0.4)
		Imagined	1.93 (0.3)	1.97 (0.4)	2.03 (0.4)

3.3. Old/new performance

Descriptive statistics for old/new performance are presented in Table 6.3. There was no main effect of stimulation condition on number of old/new errors made ($F_{(2, 68)} = 0.85$, $p = .433$). There was a main effect of error direction ($F_{(1, 34)} = 7.91$, $p = .008$), showing that participants were more likely to incorrectly classify an old item as new ($M = 33.01$, $SD = 11.8$) than to classify a new item as old ($M = 24.09$, $SD = 13.9$). Old/new performance did not interact with any variables (all ps > .08).

3.4. Effects of imagery vividness on old/new performance

Descriptive statistics for old/new vividness ratings are presented in Table 6.4. There was no main effect of stimulation condition ($F_{(2, 68)} = 0.69$, $p = .503$), indicating that stimulation did not affect vividness ratings over all items. There was a main effect of old/new response type ($F_{(1, 34)} = 18.60$, $p < .001$), showing that items correctly classified as 'old' ($M = 2.37$, $SD = 0.20$) were previously rated as higher in vividness than old items incorrectly classified as 'new' ($M = 2.28$, $SD = 0.25$). There were no interactions between any other variables (all ps > .210).

3.5. Self-report measures and task performance

Spearman's correlations between self-reported hallucination proneness (LSHS-R) and task errors indicated that neither total number of imagined words incorrectly recalled as heard ($r_s = .233$, $p = .171$), or number of heard words incorrectly recalled as imagined ($r_s = -.266$, $p = .116$) (summed across all conditions), were significantly associated with hallucination-proneness.

When broken down by condition, hallucination-proneness was positively correlated with number of imagined items recalled as heard in the temporal stimulation

condition ($r_s = .298$, $p = .039$), although this was not the case in the occipital ($r_s = .186$, $p = .279$) or sham ($r_s = .126$, $p = .465$) stimulation conditions (one-tailed). However, Hotelling's *t*-test indicated that correlation coefficients did not significantly differ between the temporal and occipital ($t(33) = 0.80$, $p = .214$), temporal and sham ($t(33) = 1.38$, $p = .089$), or occipital and sham conditions ($t(33) = 0.39$, $p = .653$). Hallucination-proneness was negatively associated with heard items being incorrectly recalled as imagined in the sham stimulation condition ($r_s = -.374$, $p = .013$), but not in the temporal ($r_s = -.105$, $p = .542$) or occipital ($r_s = -.148$, $p = .389$) conditions (one-tailed). Hotelling's *t*-test indicated that there was a significant difference between the correlation coefficients in the sham and temporal stimulation conditions ($t(33) = 1.94$, $p = .031$), and in the sham and occipital conditions ($t(33) = 1.71$, $p = .049$), but not between the temporal and occipital conditions ($t(33) = 0.28$, $p = .61$).

Table 6.3: Old/new errors (%), by interval length, error direction and stimulation condition (*M*, *SD*)

Interval length	Error direction	Stimulation condition		
		Temporal	Occipital	Sham
5 mins	Old-new	32.78 (11.4)	34.78 (12.7)	30.67 (17.1)
	New-old	22.78 (15.8)	20.56 (14.3)	20.00 (13.7)
15 mins	Old-new	35.11 (11.5)	31.6 (12.2)	33.11 (13.7)
	New-old	24.89 (15.6)	29.56 (19.8)	26.78 (12.7)

Neither retrospective estimations of vividness taken at the end of each session (see Section 2.3) ($F_{(2, 66)} = 1.46$, $p = .240$), nor estimations of the extent to which the words were imagined in the other person's voice ($F_{(2, 66)} = 0.74$, $p = .482$) differed between the three stimulation conditions.

Table 6.4: Vividness ratings for 'hear' and 'imagine' items, by interval length, response type (i.e., whether they were classified as old or new), and stimulation condition (*M*, *SD*).

Interval length	Response type	Stimulation condition		
		Temporal	Occipital	Sham
5 mins	Old	2.32 (0.3)	2.38 (0.2)	2.29 (0.3)
	New	2.22 (0.3)	2.20 (0.3)	2.22 (0.3)
15 mins	Old	2.39 (0.2)	2.42 (0.2)	2.45 (0.2)
	New	2.33 (0.2)	2.35 (0.2)	2.40 (0.3)

Analysis of the meta-cognitive estimates of task performance showed that participant estimates of the percentages of words they had classified correctly ($M = 49.67$, $SD = 11.2$) were significantly lower than the actual number of words classified correctly ($M = 54.41$, $SD = 11.2$) ($F_{(1, 33)} = 4.87$, $p = .034$). There was also a 3×2 interaction between stimulation condition (temporal/occipital/sham) and percentage of correctly classified words (actual score/meta-cognitive estimate) ($F_{(2, 66)} = 4.02$, $p = .023$). Further analysis revealed no effect of stimulation condition on total number of words

correctly classified ($F_{(2, 66)} = 0.98$, $p = .381$), but an effect of stimulation condition on the meta-cognitive estimate of correctly classified words ($F_{(2, 66)} = 4.49$, $p = .015$). Paired t-tests indicated that meta-cognitive estimates were significantly lower in the temporal stimulation condition ($M = 46.47$, $SD = 11.48$) than in the occipital condition ($M = 50.82$, $SD = 13.03$), ($t(33) = 2.51$, $p = .017$) and the sham condition ($M = 51.71$, $SD = 13.94$), ($t(33) = 2.90$, $p = .007$), but that there was no difference between estimates in the occipital and sham stimulation conditions ($t(33) = 0.43$, $p = .671$).

4. Discussion

This study used tDCS to investigate the role of left superior temporal and right anterior medial prefrontal cortices in source memory for heard and imagined verbal stimuli. Participants took part in three conditions, receiving stimulation in separate sessions to 1) right amPFC and left STG, 2) right amPFC and a left hemispheric occipital region, and 3) sham stimulation.

4.1. Reality monitoring task performance

Firstly, the key finding of this study was that there was no discernible effect of tDCS on source memory judgements. It was hypothesized that stimulation of the left STG and right amPFC, compared to the two other conditions, would lead to participants being more likely to recall words as 'heard' (corresponding to an externalising bias often shown by individuals prone to hallucinate). That stimulation in this condition did not affect task performance compared to the occipital or sham conditions suggests that the left STG may not play a crucial role in encoding source in reality

monitoring tasks. It was also hypothesized that stimulating right amPFC and an occipital control site, compared to sham stimulation, would decrease the overall number of correct source judgements. Again, results showed no difference between the conditions on task performance, implying that encoding source may not crucially rely on the right amPFC. As predicted, there was no difference between the three conditions on old/new recognition.

Previous studies into the cognitive and neuroscientific mechanisms underlying reality monitoring have mainly used neuroimaging techniques to investigate associations between levels of activation in specific regions and task performance (Mitchell & Johnson, 2009; Simons, et al., 2006; Sugimori, et al., 2014). The amPFC has been implicated in switching between stimulus-independent and stimulus-oriented cognition (Burgess, et al., 2005), variously being linked to theory of mind, meta-cognition, and episodic memory. The left STG, meanwhile, encompasses primary and secondary auditory cortex, and, in relation to reality monitoring, primary and secondary sensory cortices have been linked to imbuing imagery with perceptual detail/vividness in both the visual (Gonsalves et al., 2004; Kensinger & Schacter, 2006) and auditory (Shergill et al., 2001) domain, and/or as important in self-monitoring through a role in comparing sensorimotor predictions with incoming perceptual information (Simons, et al., 2010). The present study expanded upon previous work by utilising a technique which enables inferences into the causal role of certain brain regions in specific tasks. To our knowledge, this is the first time tDCS has been used to investigate the causal role of specific brain regions in reality monitoring.

Since stimulation of the left STG or right amPFC did not affect reality monitoring performance, it is possible that activation seen in previous neuroimaging studies is

epiphenomenal, and not crucial for successful task performance. For example, Sugimori et al. (2014) also implicated the left inferior frontal gyrus and left middle frontal gyrus in reality monitoring task performance, linking activation in these areas to perceptual vividness and cognitive operations respectively. Furthermore, Mitchell and Johnson (2009) outline a wide range of cortical and subcortical regions that have been implicated in source memory task performance, such as dorsolateral prefrontal cortex, as well as hippocampal and parahippocampal regions (which neurostimulation techniques such as used here would not be capable of directly stimulating). It is therefore possible that these regions may play a more crucial role in reality monitoring, whereas activity in the right amPFC and left STG are instead only epiphenomenally involved.

A more likely scenario is perhaps that the functional significance of amPFC and left STG is reflected in their functional connectivity with other regions, which may not have been affected by the stimulation montage used in this study. For example, whilst increasing excitability in the left STG may have secondarily affected other cortical regions, functional connectivity (correlations in levels of activity) between these regions may have remained relatively stable. Techniques that are used to introduce noise into cortical systems (e.g., transcranial random noise stimulation) may be ideal for investigating this issue. The hypothesis that functional connectivity between frontal and temporal regions may underline reality monitoring biases receives some support from a study by Wang, Metzak, and Woodward (2011), who showed that functional connectivity between amPFC and left STG was abnormal in patients with a diagnosis of schizophrenia, whilst they performed a source memory task. This idea is also consistent with studies using other task paradigms which require participants to distinguish between stimuli associated with the self or other

people. For example, Mechelli et al. (2007), using dynamic causal modelling, showed that misattribution of speech in hallucinating participants was associated with atypical connectivity between the STG and anterior cingulate cortex. Further study is needed to investigate the effects of superior temporal tDCS on functional connectivity with other regions. Future research could accomplish this by using tDCS and fMRI concurrently, which would also enable an investigation of the effects of tDCS on brain regions distal to the electrode.

One unexpected finding related to the participants' meta-cognitive estimates of their task performance. The measure of meta-cognition was included with the hypothesis that amPFC stimulation would lead to worse meta-cognitive estimates of performance (in other words, that participants would be less accurate at estimating the number of words they had correctly classified). Overall, participants underestimated the percentage of words that they had correctly classified as heard, imagined or new, but further analysis indicated that this effect was only evident in the temporal stimulation condition (compared to the occipital and sham conditions). This suggests that the effect was not due to stimulation of the right amPFC (since this was also stimulated in the occipital stimulation condition), but instead must have been due to the positioning of the anodal electrode over the left STG (or an interaction between the two electrodes). This finding seems at odds with previous research, much of which suggests that medial and lateral prefrontal areas are of crucial importance in meta-cognitive evaluation of task performance (Fleming & Dolan, 2012; Baird et al., 2013). Although impaired meta-cognition in schizophrenia has been linked to atypical activation in the superior temporal sulcus, these findings pointed towards hypoactivation (Murphy et al., 2010), rather than increased excitability which would be expected after anodal stimulation to STG, as in the

present study. These findings are, therefore, hard to interpret, and further research should attempt to replicate the present results before firm conclusions are drawn.

4.2. Perceptual vividness of imagined words

Previous research has suggested that the vividness with which internal mental events are experienced is one variable used to discriminate between internally and externally generated information; for example, in source memory paradigms, such as the one used in the present study, imagined stimuli rated as having higher perceptual detail have previously been shown to be more likely to become externally misattributed, and later recalled as 'heard' (Johnson, Raye, Wang, & Taylor, 1979; Sugimori, et al., 2014). Therefore, in the present study, we expected that words classified as heard, correctly or incorrectly, would have previously been rated as higher in vividness. Our data showed that participants rated items that were presented to 'hear' as more vivid than items that were presented to 'imagine', as might be expected. However, words that participants *later* recalled as heard (irrespective of original presentation type) were also rated as more vivid than those recalled as imagined or new. In this respect, the data provides support for previous findings, indicating that participants used information relating to perceptual vividness to make source judgements.

This finding did not interact with the type of stimulation applied, indicating that modulating excitability in the left STG or right amPFC did not affect the perceived vividness of heard or imagined words. This was not a primary hypothesis of the study; indeed, asking participants to rate perceptual vividness on a scale from 1-3 may be too coarse a measure to have been affected by a technique such as tDCS, which elicits relatively small changes in cortical excitability. However, retrospective

estimates of vividness, as well as estimates of the extent to which the words were imagined in the 'other' voice, on a larger scale (0-100), also did not differ by stimulation condition, implying that stimulation did not affect this facet of task performance.

4.3. Associations between task performance and hallucination-proneness

To our knowledge, the only study that has used neuroimaging to study the neural correlates of external misattributions on a source memory task, specifically in relation to hallucination-proneness in a non-clinical sample, was conducted by Sugimori et al. (2014), who showed that the level of activity in the left STG during external misattributions (imagine to hear errors) was positively correlated with self-reported hallucination-proneness. Therefore, we investigated whether the number of external misattributions on the source memory task was most strongly associated with hallucination-proneness when we increased excitability in the STG using anodal stimulation. Correlational analysis indeed indicated that self-reported hallucination-proneness was only significantly correlated with number of external misattributions in the temporal stimulation condition, but not the other conditions; however, the difference between the correlation coefficients did not reach significance. Interestingly, the propensity to make errors in the opposite direction (hear to imagine errors) was only negatively associated with hallucination-proneness in the sham stimulation condition. Further, this correlation coefficient was significantly different to that observed in the two active stimulation conditions. This might indicate that amPFC stimulation (the electrode position common to both active stimulation conditions) affected the tendency of hallucination-prone participants not to make internal misattributions. This finding might suggest that activity in the right prefrontal cortex may play a role, specifically in hallucination-prone individuals, in

the classification of externally generated stimulus as external. However, future research, with larger sample sizes (therefore with greater statistical power) should aim to test whether stimulation can affect performance differently in groups who report higher and lower levels of hallucination-proneness.

4.4. Limitations and future research

Although the use of tDCS allows inferences into the causal role of brain areas in specific cognitive processes, it also suffers from a number of weaknesses. The present study used $25cm^2$ electrodes, which, although typical for tDCS studies, limits the spatial resolution of the technique to relatively large cortical regions. Although there was no effect of stimulation on reality monitoring task performance, it cannot be ruled out that stimulation of multiple cortical regions under each electrode cancelled out any hypothesized effect of stimulation. For example, different regions within the medial prefrontal cortex are known to be associated with different aspects of memory, emotion and attention, at a fine spatial scale (Gilbert, Henson, & Simons, 2010). Research coupling neurostimulation and neuroimaging methods such as fMRI are needed to study the precise effects of tDCS during source memory, and could also be used to investigate potential inter-individual differences in the effects of stimulation (such as between individuals scoring high and low in proneness to hallucinations, or in clinical and non-clinical groups who experience regular hallucinations).

It should also be noted that this study only suggests that prefrontal and superior temporal regions may not be crucial in *encoding* information relating to the source of perceptual stimuli. Future research should also examine the role of these areas in the *retrieval* of source information, especially the amPFC, which has previously been

shown to be active during the retrieval of source information (Simons, Gilbert, Owen, Fletcher, & Burgess, 2005). A future study could therefore replicate the present study, but applying tDCS during the recall stage of the task. It is possible, for example, that stimulation of the left STG during retrieval could modulate the perceptual detail associated with recalled words, rather than encoded words.

4.5. Conclusions

Overall, the present study showed few discernible effects of prefrontal and/or superior temporal stimulation on source memory performance. The study replicated previous findings which have emphasized the importance of information relating to perceptual vividness in source judgements. Stimulation of the left STG was associated with reduced accuracy in estimating task performance, suggesting that participants' ability to metacognitively assess performance was impaired by increased excitability of the STG, despite no differences in reality monitoring task performance. There were some differences in the associations between self-reported hallucination-proneness and task performance between the different stimulation conditions used, although these should be treated with caution until further research specifically investigates these issues.

References

Aleman, A., Böcker, K. B. E., Hijman, R., de Haan, E. H. F., & Kahn, R. S. (2003). Cognitive basis of hallucinations in schizophrenia: role of top-down information processing. *Schizophrenia Research, 64*(2–3), 175-185.

Allen, P., Larøi, F., McGuire, P. K., & Aleman, A. (2008). The hallucinating brain: a review of structural and functional neuroimaging studies of hallucinations. *Neuroscience & Biobehavioral Reviews, 32*(1), 175-191.

Antal, A., Nitsche, M. A., Kruse, W., Kincses, T. Z., Hoffmann, K. P., & Paulus, W. (2004). Direct current stimulation over V5 enhances visuomotor coordination by improving motion perception in humans. *Journal of Cognitive Neuroscience, 16*(4), 521-527.

Baird, B., Smallwood, J., Gorgolewski, K. J. & Margulies, D. S. (2013). Medial and lateral networks in anterior prefrontal cortex support metacognitive ability for memory and perception. *Journal of Neuroscience, 33*(42), 16657-16665.

Bentall, R. P., Baker, G. A., & Havers, S. (1991). Reality monitoring and psychotic hallucinations. *British Journal of Clinical Psychology, 30*, 213-222.

Brookwell, M. L., Bentall, R. P., & Varese, F. (2013). Externalizing biases and hallucinations in source-monitoring, self-monitoring and signal detection studies: a meta-analytic review. *Psychological Medicine, 43*(12), 2465-2475.

Brunelin, J., Poulet, E., Bediou, B., Kallel, L., Dalery, J., D'Amato, T., et al. (2006). Low frequency repetitive transcranial magnetic stimulation improves source monitoring deficit in hallucinating patients with schizophrenia. *Schizophrenia Research, 81*(1), 41-45.

Buda, M., Fornito, A., Bergström, Z. M., & Simons, J. S. (2011). A specific brain structural basis for individual differences in reality monitoring. *The Journal of Neuroscience, 31*(40), 14308-14313.

Burgess, P. W., Simons, J. S., Dumontheil, I., & Gilbert, S. J. (2005). The gateway hypothesis of rostral prefrontal cortex (area 10) function. In J. Duncan, L. Phillips & P. McLeod (Eds.), *Measuring the Mind: Speed, Control and Age* (pp. 217-248). Oxford: Oxford University Press.

Finke, R. A., Johnson, M. K., & Shyi, G. C. (1988). Memory confusions for real and imagined completions of symmetrical visual patterns. *Memory and Cognition, 16*(2), 133-137.

Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1594), 1338-1349.

Ford, J. M., & Mathalon, D. H. (2005). Corollary discharge dysfunction in schizophrenia: can it explain auditory hallucinations? *International Journal of Psychophysiology, 58*(2-3), 179-189.

Frith, C. D. (1992). *The Cognitive Neuropsychology of Schizophrenia*. Hove: Lawrence Erlbaum Associates.

Gandiga, P. C., Hummel, F. C., & Cohen, L. G. (2006). Transcranial DC stimulation (tDCS): a tool for double-blind sham-controlled clinical studies in brain stimulation. *Clinical Neurophysiology, 117*(4), 845-850.

Gilbert, S. J., Henson, R. N. A., & Simons, J. S. (2010). The scale of functional specialization within human prefrontal cortex. *Journal of Neuroscience, 30*(4), 1233-1237.

Gonsalves, B., Reber, P. J., Gitelman, D. R., Parrish, T. B., Mesulam, M. M., & Paller, K. A. (2004). Neural evidence that vivid imagining can lead to false remembering. *Psychological Science, 15*(10), 655-660.

Hoffman, R. E., Wu, K., Pittman, B., Cahill, J. D., Hawkins, K. A., Fernandez, T., et al. (2013). Transcranial magnetic stimulation of Wernicke's and right homologous sites to curtail "voices": a randomized trial. *Biological Psychiatry, 73*(10), 1008-1014.

Hummel, F. C., & Cohen, L. G. (2006). Non-invasive brain stimulation: a new strategy to improve neurorehabilitation after stroke? *Lancet Neurology, 5*(8), 708-712.

Jardri, R., Pouchet, A., Pins, D., & Thomas, P. (2011). Cortical activations during auditory verbal hallucinations in schizophrenia: a coordinate-based meta-analysis. *American Journal of Psychiatry, 168*(1), 73-81.

Johnson, M. K., Foley, M. A., & Leach, K. (1988). The consequences for memory of imagining in another person's voice. *Memory and Cognition, 16*(4), 337-342.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*(1), 3-28.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*(1), 67-85.

Johnson, M. K., Raye, C. L., Wang, A. Y., & Taylor, T. H. (1979). Fact and fantasy: the roles of accuracy and variability in confusing imaginations with perceptual experiences. *Journal of Experimental Psychology: Human Learning and Memory, 5*(3), 229-240.

Jones, S. R., & Fernyhough, C. (2007). Thought as action: inner speech, self-monitoring, and auditory verbal hallucinations. *Consciousness and Cognition, 16*(2), 391-399.

Karim, A. A., Schneider, M., Lotze, M., Veit, R., Sauseng, P., Braun, C., et al. (2010). The truth about lying: inhibition of the anterior prefrontal cortex improves deceptive behavior. *Cerebral Cortex, 20*(1), 205-213.

Kensinger, E. A., & Schacter, D. L. (2006). Neural processes underlying memory attribution on a reality-monitoring task. *Cerebral Cortex, 16*(8), 1126-1133.

Lawrie, S. M., Buechel, C., Whalley, H. C., Frith, C. D., Friston, K. J., & Johnstone, E. C. (2002). Reduced frontotemporal functional connectivity in schizophrenia associated with auditory hallucinations. *Biological Psychiatry, 51*(12), 1008-1011.

McCarthy-Jones, S. R., & Fernyhough, C. (2011). The varieties of inner speech: links between quality of inner speech and psychopathological variables in a sample of young adults. *Consciousness and Cognition, 20*(4), 1586-1593.

Mechelli, A., Allen, P., Amaro, E., Fu, C. H. Y., Williams, S. C. R., Brammer, M. J., et al. (2007). Misattribution of speech and impaired connectivity in patients with auditory verbal hallucinations. *Human Brain Mapping, 28*(11), 1213-1222.

Mitchell, K. J., & Johnson, M. K. (2009). Source monitoring 15 years later: what have we learned from fMRI about the neural mechanisms of source memory? *Psychological Bulletin, 135*(4), 638-677.

Moseley, P., Fernyhough, C., & Ellison, A. (2013). Auditory verbal hallucinations as atypical inner speech monitoring, and the potential of neurostimulation as a treatment option. *Neuroscience & Biobehavioral Reviews, 37*(10), 2794-2805.

Moseley, P., Fernyhough, C., & Ellison, A. (2014). The role of the superior temporal lobe in auditory false perceptions: a transcranial direct current stimulation study. *Neuropsychologia, 62*, 202-208.

Murphy, E. R., Brent, B. K., Benton, M., Pruitt, P., Diwadkar, V., Rajarethinam, R. P., et al. (2010). Differential processing of metacognitive evaluation and the neural circuitry of the self and others in schizophrenia: a pilot study. *Schizophrenia Research, 116*(2-3), 252-258.

Nitsche, M. A., & Paulus, W. (2000). Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *The Journal of Physiology, 527*(3), 633-639.

Seal, M., Aleman, A., & McGuire, P. (2004). Compelling imagery, unanticipated speech and deceptive memory: neurocognitive models of auditory verbal hallucinations in schizophrenia. *Cognitive Neuropsychiatry, 9*(1-2), 43-72.

Shergill, S. S., Bullmore, E. T., Brammer, M. J., Williams, S. C. R., Murray, R. M., & McGuire, P. K. (2001). A functional study of auditory verbal imagery. *Psychological Medicine, 31*(2), 241-253.

Simons, C. J. P., Tracy, D. K., Sanghera, K. K., O'Daly, O., Gilleen, J., Dominguez, M., et al. (2010). Functional magnetic resonance imaging of inner speech in schizophrenia. *Biological Psychiatry, 67*(3), 232-237.

Simons, J. S., Davis, S. W., Gilbert, S. J., Frith, C. D., & Burgess, P. W. (2006). Discriminating imagined from perceived information engages brain areas implicated in schizophrenia. *NeuroImage, 32*(2), 696-703.

Simons, J. S., Gilbert, S. J., Owen, A. M., Fletcher, P. C., & Burgess, P. W. (2005). Distinct roles for lateral and medial anterior prefrontal cortex in contextual recollection. *Journal of Neurophysiology, 94*(1), 813-820.

Slotema, C. W., Blom, J. D., van Lutterveld, R., Hoek, H. W., & Sommer, I. E. (2013). Review of the efficacy of transcranial magnetic stimulation for auditory verbal hallucinations. *Biological Psychiatry, 76*(2), 101-110.

Stagg, C. J., & Nitsche, M. A. (2011). Physiological basis of transcranial direct current stimulation. *Neuroscientist, 17*(1), 37-53.

Stephane, M., Kuskowski, M., McClannahan, K., Surerus, C., & Nelson, K. (2010). Evaluation of speech misattribution bias in schizophrenia. *Psychological Medicine, 40*(5), 741-748.

Sugimori, E., Mitchell, K. J., Raye, C. L., Greene, E. J., & Johnson, M. K. (2014). Brain mechanisms underlying reality monitoring for heard and imagined words. *Psychological Science, 25*(2), 403-413.

Vinogradov, S., Luks, T. L., Simpson, G. V., Schulman, B. J., Glenn, S., & Wong, A. E. (2006). Brain activation patterns during memory of cognitive agency. *Neuroimage, 31*(2), 896-905.

Wang, L., Metzak, P. D., & Woodward, T. S. (2011). Aberrant connectivity during self–other source monitoring in schizophrenia. *Schizophrenia Research, 125*(2–3), 136-142.

Wilson, M. D. (1988). The MRC psycholinguistic database: machine readable dictionary, version 2. *Behavior Research Methods, Instruments, & Computers, 20*(1), 6-11.

Woodward, T. S., Menon, M., & Whitman, J. C. (2007). Source monitoring biases and auditory hallucinations. *Cognitive Neuropsychiatry, 12*(6), 477-494.

You, D. S., Kim, D., Chun, M. H., Jung, S. E., & Park, S. J. (2011). Cathodal transcranial direct current stimulation of the right Wernicke's area improves comprehension in subacute stroke patients. *Brain and Language, 119*(1), 1-5.

Zvyagintsev, M., Clemens, B., Chechko, N., Mathiak, K. A., Sack, A. T., & Mathiak, K. (2013). Brain networks underlying mental imagery of auditory and visual information. *European Journal of Neuroscience, 37*(9), 1421-1434.

Chapter 7

General Discussion

The previous six chapters have investigated some of the cognitive and neural mechanisms associated with hallucinations in a non-clinical sample. This chapter will provide an overview of the thesis, consider the implications of its findings and problematic issues arising from the empirical studies conducted, and discuss future avenues for research in the area.

1. Summary of findings

Chapter 1 provided a review of the most prominent cognitive neuroscientific model of auditory verbal hallucinations (AVHs), discussing the evidence that some AVHs may occur when inner speech is misattributed to an external source, due to atypical self- or source monitoring processes, and concluding that there is relatively strong evidence for this model. The review provided an outline of a number of cognitive tasks often used to assess self- or source monitoring, atypical performance on which has been associated with AVHs (see Chapter 1, Table 1.1). Recent research into the use of neurostimulation techniques as a treatment option for AVHs was also discussed, and it was argued that, if effective, neurostimulation to the left posterior superior temporal gyrus/temporoparietal junction (STG/TPJ) may affect processes involved in the monitoring of inner speech.

Chapter 3, using an auditory signal detection task, showed that auditory verbal imagery affected task performance only in hallucination-prone individuals, suggesting that they may struggle to monitor internally generated signals. Chapter 4,

using hierarchical regression, showed that a number of different variables such as inner speech phenomenology and performance on reality discrimination tasks, accounted for unique variance when predicting hallucination-proneness. Chapters 5 and 6 investigated the neural basis of performance on tasks that required participants to distinguish between self- and non-self-generated information, showing that modulating the excitability in left STG affected performance on an auditory signal detection task, but not on a reality monitoring task.

Taken as a whole, this thesis has addressed three issues. Firstly, previous research has tended to assume that a number of cognitive tasks aimed at assessing self-/source monitoring measure a single underlying cognitive mechanism, which may not be the case. Secondly, the extent to which these tasks relate to inner speech models of AVHs is questionable, and, as such, studies in this thesis have aimed to directly test inner speech or auditory imagery, and its relation to AVHs. Thirdly, very few studies have tested the cognitive neuroscientific mechanisms thought to underlie AVHs using neurostimulation techniques, limiting the types of inferences that can be made about the role of specific cortical regions, and limiting knowledge of the effects of neurostimulation on self-monitoring. This chapter will take these themes in turn, focusing on how the empirical studies in the thesis have progressed the field by addressing these issues.

2. Source monitoring, reality discrimination and self-monitoring

A number of terms are used to describe how a self-generated event, such as inner speech, may be misattributed to an external source. For example, previous studies have used tasks aimed at assessing reality monitoring (a sub-category of source

monitoring), reality discrimination, and self-monitoring. These are typically associated with performance on different tasks – reality monitoring is often tested using a source memory paradigm (Bentall, Baker, & Havers, 1991), reality discrimination using an auditory signal detection task (Aleman, Böcker, Hijman, de Haan, & Kahn, 2003), and self-monitoring using tasks requiring the participant to distinguish between the results of movements performed by themselves and by an external stimulus (Williams & Happé, 2009) (see Table 4.1, Chapter 4 for an overview of these terms). Although these tasks differ in a number of ways (whether they require online/offline decisions, whether they involve the participant in inner speech, overt speech, or a bodily movement, whether they are analysed using signal detection theory), they are often assumed to reflect one mechanism which, when dysfunctional/biased, can lead to the external misattribution of internal mental events.

The results presented in this thesis, however, do not support such a simplistic view. For example, Chapter 4 showed that, in a non-clinical sample, lower response biases on a signal detection task were associated with hallucination-proneness, whilst performance on an internal source monitoring task was not. Performance on a self-monitoring task, however, was positively associated with hallucination-proneness. This suggests that the different tasks do not index one underlying mechanism. Furthermore, Chapter 5 indicated that false alarm responses on a signal detection task were affected by modulation of excitability of the left STG, whereas Chapter 6 indicated that performance on a reality monitoring task was not affected by the level of excitability in left STG during encoding. Based on these results, it seems unlikely that the two tasks both draw on the same neural mechanisms in the superior temporal lobe (although this says nothing of potential overlap in other areas of the brain).

These studies suggest, therefore, that a more complex and multi-faceted approach is needed to explain how internal processes may become misattributed to external sources, as in AVHs. A similar point has also been made in the study of agency by Synofzik, Vosgerau, and Newen (2008), who argue that most accounts of the sense of agency do not distinguish between the 'feeling' and the 'judgement' of agency. They argue that the feeling of agency probably relates to lower level cognitive processes (including efference copy mechanisms outlined in this thesis), and is a relatively automatic process consisting of little conscious evaluation. The judgement of agency, meanwhile, relates to higher level evaluative processes, including decision-making biases and conscious evaluation. The results presented in this thesis may indicate that a similar approach would be useful when discussing the source monitoring, reality discrimination and self-monitoring literature.

One possibility is that self-monitoring tasks represent lower level mechanisms based on motor processes, such as efference copy mechanisms, whereas measures such as the jumbled speech task used in Chapter 4 represent higher level decision-making biases which are not necessarily linked to the motor system. This would be consistent with the findings in Chapter 4, which showed that self-monitoring performance was *positively* associated with hallucination-proneness. As discussed, it is possible that a bias in higher level decision-making processes may be associated with non-clinical hallucination-proneness, but this only develops into the more frequent experiences associated with psychosis when the self-monitoring/efference copy system is also dysfunctional. It is unclear, however, whether reality discrimination tasks such as the auditory signal detection task used in this thesis draw on these higher or lower level processes (or both). This could be investigated by testing the involvement of lower level motor processes in reality discrimination

tasks. Such a study could, for example, test whether reality discrimination biases are affected by arresting motor activity, by requiring the participant to complete a signal detection task that requires simultaneous use of the mouth (such as chewing gum, or holding a pencil in the mouth; Reisberg et al., 1989). If such a manipulation affected response biases on this task, this would indicate that low level motor processes are involved in false perceptions in signal detection. If not, this might indicate that higher level decision making biases underlie differential task performance in hallucination-prone individuals.

3. Inner speech models of AVH

Another problematic aspect of previous research into AVHs is that it often purports to be testing inner speech models, despite the fact that the tasks used do not engage the participant in inner speech or auditory verbal imagery. As such, the assumption is made that, for example, speaking aloud uses the same or similar cognitive and neural mechanisms as inner speech. The studies in this thesis have all engaged the participant in auditory verbal imagery (Chapters 3 and 6), assessed participants' experience of inner speech (Chapter 4), or stimulated cortical regions that have previously been shown to be involved in inner speech (Chapters 5 and 6). The results from these studies provide evidence that 1) a tendency to use dialogic inner speech is associated with hallucination-proneness (consistent with the findings of McCarthy-Jones et al., 2011), 2) use of auditory verbal imagery leads to a lower response bias on signal detection tasks in hallucination-prone individuals, and 3) stimulating left STG, an area previously shown to be overactive in individuals with a diagnosis of

schizophrenia during inner speech (Simons et al., 2010) leads to a higher rate of false perceptions in a signal detection task.

However, a relatively small amount of research has investigated the precise cognitive or neural mechanisms involved in false perceptions on auditory signal detection paradigms. One study, carried out by Vercammen and Aleman (2010), showed that semantic priming increased the number of false perceptions in a signal detection task. The authors suggested that this reflected aberrant top-down processing in hallucination-prone individuals, in which prior expectations were too highly weighted. This is consistent with the findings presented in Chapter 3 (Experiment 2), which showed that cuing hallucination-prone participants with a sentence on a signal detection task led to a lower response bias. However, our findings indicated that rather than simply being an effect of semantic expectation, this was dependent on the concurrent use of auditory verbal imagery. Therefore, this finding is equally consistent with the explanation that false alarms on auditory signal detection tasks reflect external misattributions of internal mental imagery. Alternatively, it is possible that these seemingly different types of explanation can be subsumed under a broader framework of predictive coding (see Wilkinson, 2014, and below).

The findings in this thesis not only provide support for inner speech models of AVHs, using tasks with more relevance to inner speech, but also highlight that it is possible to study inner experience in an experimental setting. This opens the door to numerous options for future studies. One outstanding question for inner speech models is why some instances of inner speech may become misattributed, but not others. Building on the findings of Chapter 3, this could be addressed by manipulating the type of inner speech that participants are asked to use (e.g., dialogic vs. monologic, first/second/third person, in another's voice vs. one's own voice)

whilst completing signal detection tasks, and observing the extent to which each type

had an effect on response biases. Based on the studies investigating the

phenomenology of AVHs (McCarthy-Jones et al., 2012; Nayani & David, 1996), it

might be predicted that inner speech in another's voice, of a dialogic quality, and in

the second or third person, would be most likely to become misattributed.

4. Neurostimulation and AVHs

A key issue raised in Chapter 1 related to the use of neurostimulation as a treatment

option for AVHs. Chapter 5 indicated that modulating the level of excitability of the

left posterior STG using cathodal transcranial direct current stimulation (tDCS)

affected the number of false alarm responses on an auditory signal detection task

Since the publication of Chapters 1 and 5, new evidence has been published

suggesting that cathodal tDCS to the left STG may reduce AVH frequency, with a

concurrent improvement in source monitoring ability (Mondino, Haesebaert, Poulet,

Suaud-Chagny & Brunelin, in press), and also affect the N1 ERP response to self-

produced vocalisations (thought to be related to efference copy mechanisms)

(Nawani et al., 2014). These findings provide support for one argument made in

Chapter 1: that the therapeutic effect of neurostimulation may result from the role of

the left STG in monitoring inner speech. However, the findings of Mondino et al. are

not consistent with the findings presented in Chapter 4 or 6. This is because the task

they used was an *internal* source monitoring task, in which participants receiving

daily tDCS were asked to complete a task in which they had to recall whether they

had spoken or imagined a word. The data presented in Chapter 4 suggested that this

type of source monitoring task was not associated with proneness to hallucinations.

Furthermore, Chapter 6 suggested that performance on reality monitoring tasks (requiring the participant to recall whether they imagined or heard a word) were not affected by modulating excitability of the left STG. One possibility is that, rather than reflecting an underlying mechanism of AVHs, biased performance on internal source monitoring tasks actually results from the frequent voice-hearing experiences associated with a diagnosis of schizophrenia, and that the therapeutic effect of tDCS of reducing the frequency of AVHs in fact causes the improvement in internal source monitoring (rather than vice versa). Another possibility is that discontinuities between hallucinatory experiences assessed by the self-report scales used in this thesis and those experienced by patients with a diagnosis of schizophrenia (or other psychotic disorders) are reflected in differences in performance on internal source monitoring. Future studies should aim to compare performance on this type of task between patient groups that do or do not hallucinate, and non-clinical samples that report differing levels of hallucination-proneness.

5. Theoretical and conceptual implications

One framework that may be useful in explaining such cognitive mechanisms is the predictive coding framework (PCF, sometimes also referred to as the 'predictive processing framework', or PPF), a Bayesian approach that assumes that much (if not all) of what the brain does is aimed at predicting subsequent incoming information and then minimising the error associated with these predictions (see Clark, 2013, and Hohwy, 2014, for reviews). Two papers have specifically attempted to use a PCF approach to explain AVHs (and other positive symptoms of schizophrenia). Fletcher and Frith (2009) link excessive dopaminergic activity to atypical precision weighting

(that is, how much weight prediction errors are given in shaping further predictions), speculating that this may lead to the formation and persistence of delusional beliefs, as well as AVHs. More recently, Wilkinson (2014) has applied the PCF to specific models of AVHs, including inner speech models. The PCF might predict that the main difference between self- and non-self-generated actions would be higher precision weighting attached to error associated with self-generated actions (since the predictions should be more precise). Activation in sensory cortical areas would typically reflect prediction error – hence, activation in these areas is typically attenuated in response to self-generated actions. Atypical weighting of prediction error could therefore lead to high levels of sensory cortical activity (i.e., a small amount of error may be highly weighted), and potentially the misattribution (or misperception) of inner speech.

There are two main implications of this approach to the work discussed in this thesis. Firstly, since two studies modulated levels of activity in secondary auditory cortical regions using tDCS, it is possible that, viewed through the PCF, this would have the effect of increasing the error on sensory predictions. For example, increasing excitability in the left STG could lead to a higher level of prediction error on incoming sensory information, passed up the hierarchy, with the result being that subsequent predictions are altered (for example, regarding the presence or absence of a stimuli, as found in Chapter 5). A second implication of this approach is that it does not necessarily privilege prediction of the sensory consequences of motor actions, since the brain must predict all incoming sensory information. It is therefore possible that efference copy models of AVHs (and of self-monitoring in general) can be subsumed under the more general PCF, and findings which suggest externalising biases are associated with AVHs, even in the absence of self-produced speech (Allen

et al., 2004) can be explained in terms of atypical error weighting. The previously mentioned study by Vercammen and Aleman (2010), in which semantic expectations were found to modulate auditory false perceptions in noise in hallucination-prone individuals, can also be explained with reference to high precision weighting. Undoubtedly, linking AVHs, inner speech and self-monitoring to the PCF is an area ripe for investigation. In particular, exploring the extent to which involvement of the motor system is necessary for inner speech and AVHs will help to distinguish between traditional self-monitoring and source monitoring theories and a broader predictive coding model. On a neural basis, using functional neuroimaging to investigate activity in sensory cortices to errors in prediction of both self-generated and non-self-generated stimuli, in both hallucinating and non-hallucinating individuals, would be one way to test the ideas stemming from the PCF, outlined above.

6. Limitations and future research

The research presented in this thesis has a number of limitations. Some of these have already been covered (e.g., limitations of tDCS as a technique, such as lack of spatial resolution) and so will not be repeated in this section. One limitation which is worthy of discussion, though, relates to the use of a non-clinical sample throughout. The research presented has relied heavily on self-report measures of the experiences of hallucinations in mainly student samples. Although this has some advantages (lack of anti-psychotic medication and other symptoms of psychosis which may confound results, as well as the relative ease with which larger samples can be recruited), it inevitably brings some limitations to the conclusions that can be made. Johns et al.

(2014) have recently explored issues relating to AVHs in persons with and without a need for care. They argue that future research should be careful to distinguish between individuals that are hallucination-prone (i.e., score highly on self-report measures such as used in this thesis), healthy voice-hearers (who experience frequent voice-hearing experiences, often assessed by semi-structured interviews, but are not distressed or impaired by their experiences; Sommer et al., 2010) and clinical voice-hearers (often with a diagnosis of schizophrenia or other psychotic disorder). They also suggest that it is useful to distinguish between a continuum of experience (between, for example, intrusive mental imagery and more frequent AVHs) and a continuum of risk of developing problematic AVH with a need for care. The results presented in this thesis provide evidence that, at least with regard to hallucination-proneness assessed via self-report, the continuum of experience may be linked to specific cognitive mechanisms, although they say less about the risk of such experiences developing into problematic psychotic experiences.

Future research should also attempt to link cognitive mechanisms to potential discontinuities in experience. For example, an increasingly held view is that there may be qualitatively different subtypes of AVHs, which may be associated with different underlying cognitive mechanisms (Jones, 2010; McCarthy-Jones et al., 2014). Inner speech based AVHs, for example, may be based on biased monitoring processes, whereas hallucinations best described as 'hypervigilance' to the external environment (Dodgson & Gordon, 2009; Garwood, Dodgson, Bruce, & McCarthy-Jones, 2013) could be better explained as a result of biased attentional processes. Future research should attempt to explore the extent to which different cognitive mechanisms can explain continuities and discontinuities in hallucinatory experiences.

As such, future research should aim to apply similar methods to hallucinating and non-hallucinating individuals, for example with diagnoses of schizophrenia. In particular, investigating performance on the tasks used in Chapter 4 in a large clinical sample, to determine the extent to which they are independently associated with specific symptoms or experiences, should be a priority. It would also be predicted that the use of AVI during signal detection (Chapter 3) would have a larger effect on hallucinating patients than on non-hallucinating patients, and that the therapeutic effect of neurostimulation to the left posterior STG may be associated with less biased performance on auditory signal detection tasks.

Another limitation relates to the use of signal detection theory (SDT) to analyse task performance. Although SDT is crucial in distinguishing between patterns of performance due to response bias and differing sensitivity, it relies on a number of assumptions, which cannot be directly tested within the current data. For example, SDT calculations assume normality of the signal and noise distributions, as well as equality of signal and noise standard deviations; to test these assumptions, a task is required which allows participants to give confidence ratings on each trial. Most previous studies which apply SDT to the study of hallucinations have not used this measure, with a notable exception being Vercammen, de Haan, and Aleman (2008). However, these authors do not present any analyses of whether the SDT assumptions were met. Future research should address this issue by using a task that asks the participants for confidence ratings after each trial.

One clear avenue for future research would also be to combine the relative strengths of neurostimulation and neuroimaging techniques such as fMRI. For example, a study which concurrently used tDCS and fMRI whilst the participant performed an auditory signal detection task would allow inferences to be made regarding cortical

regions (and the interaction between them) other than those directly underneath the stimulating electrodes. It might be predicted that, as well as changes in activity levels in the left posterior STG, primary auditory cortical regions in the anterior STG, and perhaps anterior cingulate regions implicated in self-/reality monitoring, would be affected. An important analysis would also be to investigate the possible effects of tDCS on functional connectivity, particularly between frontal and temporal regions, during both auditory signal detection and source memory. Similarly, combining tDCS and EEG would allow investigation into the effect of left STG stimulation on event-related potentials (such as the N1 response thought to be associated with the efference copy mechanism), as well as cortical oscillations which may reflect synchronous activity between brain regions (Ford, Roach, Faustman, & Mathalon, 2008).

Finally, a relatively new technique, transcranial alternating current stimulation (tACS), may be a promising tool to investigate the role of cortical oscillations in cognitive functions (Ali, Sellers, & Fröhlich, 2013). tACS is capable of enhancing or interfering with the frequency of cortical oscillations (Antal & Paulus, 2013), thereby allowing inferences into the importance of certain oscillatory patterns in specific cognitive functions. This is of particular interest to the study of AVHs, since it has been hypothesized that typical efference copy/corollary discharge function may be reflected in synchronized oscillatory activity between cortical regions. For example, Ford et al. (2008) showed that synchronous oscillatory activity in the beta and gamma bands over sensory-motor cortex was lower in patients with a diagnosis of schizophrenia (not specific to those who hallucinated) before a self-generated action, compared to healthy controls. Gamma band synchrony was interpreted as reflecting efference copy signals, which were dysfunctional in the patient group.

Future research could therefore test this hypothesis further by selectively entraining or interfering with oscillatory activity to test for an effect on self-monitoring. It would be predicted that enhancing gamma band synchrony in motor cortical regions using tACS would improve self-monitoring performance (whereas interfering may have the opposite effect).

7. Conclusions

The four experimental chapters in this thesis have provided evidence that a number of variables, relating to inner speech usage and phenomenology, reality discrimination and self-monitoring, are associated with proneness to hallucinations in a non-clinical sample. Although previous research has tended to assume that different tasks are related to one underlying process, results have generally indicated that this would be an over-simplification and may not be the case. Two experimental studies also modulated activity in frontal and temporal cortical regions using neurostimulation. These studies showed that increasing excitability in the left posterior superior temporal gyrus affected the number of auditory false perceptions on a signal detection (reality discrimination) task, but did not affect performance on a source memory (reality monitoring) task. Areas for future research are outlined, including using a wider variety of cognitive neuroscientific techniques to study inner speech and self-/reality monitoring, as well as extending the use of the paradigms to clinical populations.

References

Aleman, A., Böcker, K. B. E., Hijman, R., de Haan, E. H. F., & Kahn, R. S. (2003). Cognitive basis of hallucinations in schizophrenia: role of top-down information processing. *Schizophrenia Research, 64*(2-3), 175-185.

Ali, M. M., Sellers, K. K., & Fröhlich, F. (2013). Transcranial alternating current stimulation modulates large-scale cortical network activity by network resonance. *The Journal of Neuroscience, 33*(27), 11262-11275.

Allen, P., Johns, L. C., Fu, C. H. Y., Broome, M. R., Vythelingum, G. N., & McGuire, P. K. (2004). Misattribution of external speech in patients with hallucinations and delusions. *Schizophrenia Research, 69*(2–3), 277-287.

Antal, A., & Paulus, W. (2013). Transcranial alternating current stimulation (tACS). *Frontiers in Human Neuroscience, 7*.

Bentall, R. P., Baker, G. A., & Havers, S. (1991). Reality monitoring and psychotic hallucinations. *British Journal of Clinical Psychology, 30*, 213-222.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 181-204.

Dodgson, G. & Gordon, S. (2009). Avoiding false negatives: are some auditory hallucinations an evolved design flaw? *Behavioural and Cognitive Psychotherapy, 37*(3), 325-334.

Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience, 10*(1), 48-58.

Ford, J. M., Roach, B. J., Faustman, W. O., & Mathalon, D. H. (2008). Out-of-synch and out-of-sorts: dysfunction of motor-sensory communication in schizophrenia. *Biological Psychiatry, 63*(8), 736-743.

Garwood, L., Dodgson, G., Bruce, V., & McCarthy-Jones, S. (2013). A preliminary investigation into the existence of a hypervigilance subtype of auditory hallucination in people with psychosis. *Behavioural and Cognitive Psychotherapy, FirstView*, 1-11.

Hohwy, J. (2014). *The Predictive Mind*. Oxford: Oxford University Press.

Johns, L. C., Kompus, K., Connell, M., Humpston, C., Lincoln, T. M., Longden, E. et al. (2014). Auditory verbal hallucinations in persons with and without a need for care. *Schizophrenia Bulletin, 40*(S4), S255-S264.

Jones, S. R. (2010). Do we need multiple models of auditory verbal hallucinations? Examining the phenomenological fit of cognitive and neurological models. *Schizophrenia Bulletin, 36*(3), 566-575.

McCarthy-Jones, S. R. & Fernyhough, C. (2011). The varieties of inner speech: links between quality of inner speech and psychopathological variables in a sample of young adults. *Consciousness and Cognition, 20*(4), 1586-1593.

McCarthy-Jones, S. R., Thomas, N., Strauss, C., Dodgson, G., Jones, N., Woods, A. et al. (2014). Better than mermaids and stray dogs? Subtyping auditory verbal hallucinations and its implications for research and practice. *Schizophrenia Bulletin, 40*(S4), S275-S284.

McCarthy-Jones, S. R., Trauer, T., Mackinnon, A., Sims, E., Thomas, N., & Copolov, D. L. (2012). A new phenomenological survey of auditory hallucinations: evidence for subtypes and implications for theory and practice. *Schizophrenia Bulletin, 40*(1), 231-235.

226

Mondino, M., Haesebaert, F., Poulet, E., Suaud-Chagny, M., & Brunelin, J. (in press). Fronto-temporal transcranial direct current stimulation (tDCS) reduces source-monitoring deficits and auditory hallucinations in patients with schizophrenia. *Schizophrenia Research,* dx.doi.org/10.1016/j.schres.2014.10.054.

Nawani, H., Bose, A., Agarwal, S. M., Shivakumar, V., Chhabra, H., Subramaniam, A., et al. (2014). Modulation of corollary discharge dysfunction in schizophrenia by tDCS: preliminary evidence. *Brain Stimulation, 7*(3), 486-488.

Nayani, T. H. & David, A. S. (1996). The auditory hallucination: a phenomenological survey. *Psychological Medicine, 26*(1), 177-189.

Reisberg, D., Smith, J. D., Baxter, D. A., & Sonenshine, M. (1989). "Enacted" auditory images are ambiguous; "pure" auditory images are not. *Quarterly Journal of Experimental Psychology, 41*(3), 619-641.

Simons, C. J. P., Tracy, D. K., Sanghera, K. K., O'Daly, O., Gilleen, J., Dominguez, M.-d.-G., et al. (2010). Functional magnetic resonance imaging of inner speech in schizophrenia. *Biological Psychiatry, 67*(3), 232-237.

Sommer, I. E. C., Daalman, K., Rietkerk, T., Diederen, K. M., Bakker, S., Wijkstra, J., & Boks, M. P. M. (2010). Healthy individuals with auditory verbal hallucinations; who are they? Psychiatric assessments of a selected sample of 103 subjects. *Schizophrenia Bulletin, 36*(3), 633-641.

Synofzik, M., Gottfried, V. & Newen, A. (2008). Beyond the comparator model: a multifactorial two-step account of agency. *Consciousness and Cognition, 17*(1), 219-239.

Vercammen, A., & Aleman, A. (2010). Semantic expectations can induce false perceptions in hallucination-prone individuals. *Schizophrenia Bulletin, 36*(1), 151-156.

Vercammen, A., de Haan, E. H. F., & Aleman, A. (2008). Hearing a voice in the noise: auditory hallucinations and speech perception. *Psychological Medicine, 38*(8), 1177-1184.

Wilkinson, S. (2014). Accounting for the phenomenology and varieties of auditory verbal hallucination within a predictive processing framework. *Consciousness and Cognition, 30*, 142-155.

Williams, D. & Happé, F. (2009). Pre-conceptual aspects of self-awareness in autism spectrum disorder: the case of action-monitoring. *Journal of Autism and Developmental Disorders, 39*(2), 251-259.

Appendix 1 – eligibility criteria for neurostimulation

Cognitive Neuroscience Research Unit
Wolfson Research institute
University of Durham, Queen's Campus   **tel: 0191 334 0430**
University Boulevard   **fax: 0191 334 0452**
Thornaby   **amanda.ellison@durham.ac.uk**
Stockton-on-Tees
TS17 6BH

Subject Questionnaire

If you agree to take part in this study, please answer the following questions. The information you provide is for screening purposes only and will be kept completely confidential.

Have you ever suffered from any neurological or psychiatric conditions?   YES/NO
If YES please give details (nature of condition, duration, current medication, etc).

Have you ever suffered from epilepsy, febrile convulsions in infancy or had recurrent fainting
spells?   YES/NO

Does anyone in your immediate or distant family suffer from epilepsy?   YES/NO
If YES please state your relationship to the affected family member.

Do you suffer from any skin allergies e.g. rash, eczema?   YES/NO

Have you ever undergone a neurosurgical procedure (including eye surgery)?   YES/NO
If YES please give details.

Do you currently have any of the following fitted to your body?   YES/NO
i)   Heart pacemaker
ii)   Cochlear implant
iii)   Medication pump
iv)   Surgical clips

Are you currently taking any unprescribed or prescribed medication, including anti-malarials.
If YES please give details.   YES/NO

Are you left or right handed?   LEFT/RIGHT

Subject Consent
I (please give full name in CAPITALS)_____confirm that I have read the
letter of invitation and have completed the above questionnaire. I confirm that I am not taking
recreational drugs and have not participated in a TMS/tDCS/tACS experiment already today and feel
well rested. The nature, purpose and possible consequence of the procedures involved have been
explained. I understand that I may withdraw from the study at any time.

Signature_____   Date_____