Schizophrenia Bulletin Open
https://doi.org/10.1093/schizbullopen/sgac050

Real-Time Symptom Capture of Hallucinations in Schizophrenia with fMRI: Absence of Duration-Dependent Activity

Karanvir Gill[1,2], Chantal Percival[1,2], Meighen Roes[1,2], Leo Arreaza[1], Abhijit Chinchani[1,2], Nicole Sanford[1,2,•], Walter Sena[3,•], Homa Mohammadsadeghi[4], Mahesh Menon[2], Matthew Hughes[5], Sean Carruthers[5,•], Philip Sumner[5], Will Woods[5], Renaud Jardri[7,•], Iris E. Sommer[8], Susan L. Rossell[5,6], and Todd S. Woodward[1,2,*,•]

[1]BC Mental Health and Addictions Research Institute, Vancouver, BC, Canada; [2]Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada; [3]Instituto de Psiquiatria, Universidade Federal do Rio de Janeiro, Av. Venceslau Braz, 71, Rio de Janeiro, RJ, Brazil; [4]Iran University of Medical Sciences, Tehran, Iran; [5]Centre for Mental Health, School of Health Sciences, Swinburne University, Melbourne, Australia; [6]St Vincent's Mental Health, St Vincent's Hospital, Melbourne, Australia; [7] Univ Lille, INSERM U-1172, CHU Lille, Lille Neuroscience and Cognition Centre, Plasticity & SubjectivitY Team, Lille, France; [8]Department of Neuroscience, University Medical Center Groningen, Groningen, The Netherlands

*To whom correspondence should be addressed; Room A3-A116, BC Mental Health & Addictions Research Institute – Translational Research Building, 3rd Floor, 938 W. 28th Avenue, Vancouver, British Columbia, Canada, V5Z 4H4; tel: 604-875-2000x4724, fax: 604-875-3871, e-mail: Todd.S.Woodward@gmail.com

Background: **While advances in the field of functional magnetic resonance imaging (fMRI) provide new opportunities to study brain networks underlying the experience of hallucinations in psychosis, there are methodological challenges unique to symptom-capture studies.** ***Study Design:*** **We extracted brain networks activated during hallucination-capture for schizophrenia patients when fMRI data collected from two sites was merged (combined $N = 27$). A multidimensional analysis technique was applied, which would allow separation of brain networks involved in the hallucinatory experience itself from those involved in the motor response of indicating the beginning and end of the perceived hallucinatory experience. To avoid reverse inference when attributing a function (e.g., a hallucination) to anatomical regions, it was required that longer hallucinatory experiences produce extended brain responses relative to shorter.** ***Study Results:*** **For radio-speech sound files, an auditory perception brain network emerged, and displayed speech-duration-dependent hemodynamic responses (HDRs). However, in the hallucination-capture blocks, no network showed hallucination-duration-dependent HDRs, but a retrieved network that was anatomically classified as motor response emerged.** ***Conclusions:*** **During symptom capture of hallucinations during fMRI, no HDR showed duration dependence, but a brain network anatomically matching the motor response network was retrieved. Previous reports on brain networks detected by fMRI during hallucination capture are reviewed in this context; namely, that the brain networks interpreted as involved in hallucinations may in fact be involved only in the motor response indicating the onset of the hallucination.**

Key words: hallucinations fMRI psychosis/functional brain networks/speech perception

Introduction

Auditory verbal hallucinations (AVHs) are speech perceptions that occur in the absence of an external stimulus and are primary symptoms of psychosis, such that 60%–80% of people with schizophrenia spectrum disorder report experiencing them.[1,2] The propensity to hallucinate has been linked with neural hyperactivity in voice-selective regions of the superior temporal gyrus (STG),[3–7] and these voice-selective cortical regions have been reported to activate in symptom capture studies on hallucinations.[8,9] Repeated transcranial magnetic stimulation (rTMS) of voice-selective cortical regions was also reported to reduce the intensity of hallucinations.[10–12] Therefore, much of the AVH literature to date has taken a region-of-interest based approach focusing on voice-selective regions of the STG.

AVHs are unlikely to arise exclusively from hyperactivity in the STG. For example, the breakaway speech/unbidden thoughts account of hallucinations[13–15] puts forward that AVH may occur when self-monitoring breaks down, possibly due to reduced activation in the dorsal anterior cingulate cortex (dACC)/supplementary

Table 1. Assessments of Spatial, Temporal and Experimental Validity, as a Function of Retrieved Networks.

Data	Experiment	Network	Spatial Validity (Fisher's *z* score match to task-based network template)	Temporal Validity (Time effect size: η_p^2)	Experimental Validity (Duration × Time interaction effect size: η_p^2)
Melbourne	Radio Speech (S/M/L)	C2: Auditory Perception	✓(1.04)	✓(.65)	✓(.22)
		C3: Focus on Visual Features	†(.64)	✓(.24)	†(.05)
		C1	n/a	✗	n/a
Merged: Melbourne	Hallucinations (S/L)	C1	n/a	✗	n/a
		C2: Response	✓(.88‡)	✓(.17)	✗
		C3: Focus on Visual Features	†(.48)	✓(.27)	✗
Merged: Utrecht	Hallucinations (S/L)	C2: Response	✓(.88‡)	✓(.38)	✗
		C1, 3	n/a	✗	n/a

Note. ✓ indicates clear pass with a large effect (η_p^2 >.15; Z >.70), and
† indicates a marginal pass with a small effect (η_p^2<.15; *z* <.80). ✗indicates conditions not met. n/a indicates that the cell is irrelevant due to not meeting criteria at the level of temporal validity.
‡ indicates same anatomical network in merged analysis.

motor areas (SMA) during the hallucinatory experience.[16] Metacognitive or belief-based influences are also likely play a role[17–19]; therefore, a network-based approach is important for investigating the biological underpinnings of hallucinations.

In functional magnetic resonance imaging (fMRI) symptom capture studies of hallucinations, activity in the STG is typically reported, as are a number of other language based regions, including Broca's area, anterior insula, precentral gyrus, frontal operculum, inferior parietal lobule, hippocampus, parahippocampal regions and in motor areas such as the inferior frontal gyrus, cerebellum, insula, and postcentral gyrus.[20–22] However, in most symptom capture studies, the experimental procedure to monitor hallucinations in the scanner consists of participants pressing a button or squeezing a ball, to indicate the onset and offset of hallucinations,[20–22] with exceptions being relatively rare e.g.,.[23–25] Since fMRI measures the blood-oxygen-level-dependent (BOLD) signal increases in response to cognitive events, the timing of hallucinations onset/offset is confounded with response, leading to complications separating the sensorimotor (response) network with a potential network underlying hallucination.

Exacerbating this experimental/timing confound of hallucinations events with response events is that many analysis methods for task-based fMRI interpret the significance of beta weights (or the significance of differences of beta weights) derived from regressing the BOLD signal onto an assumed and synthetic model of the hemodynamic response (HDR) shape.[26] Even if this synthetic HDR shape is adjusted for the reported duration of hallucination events, it restricts interpreted results exclusively to BOLD signal changes that conform to an assumed model of the HDR, and both the HDR resulting from the AVH, *and* the response process to report the AVH, could partially match the synthetic HDR shape model, resulting in conflation of multiple cognitive operations and their underlying brain networks on the resulting statistical parametric maps.

What results from this process is a brain activation map, interpreted as reflecting hallucinations when it may in fact be reflecting the response process used to indicate to the experimenter the presence of a hallucination, or any other cognitive process occurring at the time of the response, such as preparing a response, inspection of an internal representation, or the metacognitive process of becoming aware of an internal representation that requires a response. Making conclusions about the functions attributable to an anatomical region based on observed activation alone is known as the fallacy of reverse inference. A direct linkage of an anatomical pattern of brain activity to a cognitive process is only on solid ground if there is a one-to-one mapping between the anatomical region one hand, and the proposed cognitive operation (e.g., hallucination) on the other.[27] So far, one-to-one mapping between anatomy and cognitive operations such as hallucinations has not been possible, because there is no anatomical representation that is specific to hallucinations, and thus they cannot be directly linked.[27–31]

To counter (1) the experimental/timing confound of AVH events with response events, (2) mutual conflation of AVH and response events due to both partially matching the synthetic HDR shape model, and (3) the fallacy of reverse inference (i.e., linking an anatomical region to a cognitive process in the absence of additional information), fundamental changes in the methods applied to detect the brain networks elicited by event timing in during hallucination-capture fMRI are required. First, instead

of restricting results to partial matches to an assumed HDR displayed on brain images, an assumption-free finite impulse response (FIR) model can be used, which allows any event-related HDRs shape signal to emerge. Second, dimensional analysis methods such as principal component analysis (PCA) allow anatomical dimensions to separate instead of being merged onto one image. Third, observation of experiment-induced changes in the HDR provide additional information that is more interpretable than anatomical information alone, and can help to reduce the risk of committing the reverse-inference fallacy by providing empirical evidence for attaching a cognitive process to an anatomical depiction. Namely, the nature of the HDR shape should change as the nature of the hypothesized cognitive operations change. In the case of the current study, the duration of the HDR should increase with the duration of the radio speech event or experienced hallucination, just as is reliably observed with short/long durations of memory maintenance in working memory.[32–35]

Constrained principal component analysis for fMRI, fMRI-CPCA,[33,35] provides dimensional representation of brain networks captured using a FIR model for AVH timing. Retrieved networks can be anatomically compared templates of previously derived networks which have known anatomical configurations and associated cognitive functions, documented through inspection of network- and task-condition-specific HDRs over a wide range of tasks, and these include response and auditory perception.[36,37] This allows direct observation of the duration of detected cognitive events for each network separately, without requiring the assumptions/models of the assumed HDR shape that are typically used, as described above.

In addition to analysis of FIR-model predictable BOLD signal variance with PCA, we built experimental manipulations into the study to aid with separation of response from auditory perception networks, and to reduce the risk of misplaced reverse inference. Specifically, to provide evidence that hallucination-driven HDR have been recorded by fMRI, the following validity requirements were set: (1) spatial validity, (2) temporal validity, and (3) experimental validity.[38] Spatial validity requires observation of known network configurations.[36,37,39–41] For example, the sensorimotor (response) network is expected to involve brain regions such as the bilateral supplementary motor area (SMA), dorsal anterior cingulate cortex (dACC), and insula, as well as left somatomotor areas and right cerebellum for a one-handed response[33] Figure 7,[36] but the auditory perception network is expected to be dominated by the superior temporal gyrus[36,42] Component 7, Figure S3. For temporal validity, a biologically plausible HDR shape must be associated with an anatomically valid network to ensure that BOLD signal is likely being detected. For experimental validity, the nature of the HDR shape should change as the nature of the invoked cognitive operations change. In the case of the current study, the duration of the HDR should increase with the duration of the experienced hallucination.

Symptom capture data was analyzed by merging separate datasets from two sites (Melbourne and Utrecht), and radio speech events were also collected at the Melbourne site only. Our approach was to test spatial, temporal and experimental validity in the external (radio) sound timing from the Melbourne data (i.e., short/medium/long durations for experimental validity), and for the hallucinatory experiences in a dataset merged from the two sites (i.e., short/long durations for experimental validity). Temporally, we expected to see a pattern of increased HDR from baseline to peak, with the HDR duration at peak level determined by duration of radio speech/hallucination, before returning back to baseline. By utilizing data from two sites we are able to increase sample sizes so that it was possible to collect together a greater range in both frequency and duration of hallucinations, facilitating identification of these networks.

Methods and Materials

Participants

Melbourne. Seventeen schizophrenia patients and thirty-one healthy control participants were included in the analysis of the radio speech stimuli, and twelve of those schizophrenia patients also contributed data to the symptom capture study completed in Melbourne. Supplementary table S1 provides the demographic information of the participants and the scores on the Positive and Negative Syndrome Scale (PANSS) for the schizophrenia patients.

Utrecht. Fifteen schizophrenia patients were included in the analysis of the hallucinations from the symptom capture study completed in Utrecht. Supplementary table S1 provides the demographic information of the participants and their scores on the PANSS. These 15 patients are a subset of a sample of 19 patients previously compared to a sample of nonpsychotic voice hearers using a different analysis method in published work.[43]

Tasks

Melbourne. The task completed by participants from the Melbourne site involved indicating the start and end of (i) radio speech clips, and (ii) experienced hallucinations, using a dominant hand button-press response. For more details, see Supplementary Material.

Utrecht. The task completed by participants from the Utrecht site was to indicate the beginning and end of

hallucinations using a dominant hand balloon squeeze and release. For more details, see Supplementary Material.

Image Acquisition

Specifics about the fMRI parameters and preprocessing of functional scans are given in the Supplementary Material.

Timing

Details of the various fMRI-CPCA and repeated measures ANOVA analyses conducted with different samples of participants from both sites can be found in the Supplementary Material.

Data Analysis

Data Analysis was carried out using fMRI-CPCA,[33,35] as described in detail in the Supplementary Material. Table 1 provides assessments of spatial, temporal and experimental validity, as a function of all retrieved networks.

Results

Melbourne Radio Speech (Short/Medium/Long)

Three components were extracted for the Melbourne radio speech experiment, as determined by examining the scree plot.[44,45] Component 1 did not retrieve a biologically plausible HDR shape, so is reported only in the Supplementary Material (supplementary sigure S4A/B).

Component 2: auditory perception network. The anatomical regions associated with Component 2 are outlined in figure 1A and the anatomical description of component two is presented in supplementary table S5. The anatomical pattern matched well to that in the auditory perception network Fisher's $z = 1.04$;[36] [AUD,42] Component 7, Figure S3, with bilateral peaks in right superior temporal gyrus (xyz: 60, −16, −2), and left planum temporale (xyz: −57 −19, 1).

Figure 1B displays the estimated HDR shape for Component 2. Component 2 displayed a biologically plausible HDR, and a highly significant main effects Time, $F(19, 874) = 84.31$, $P < .001$, $\eta_p^2 = .65$, clearly meeting the requirement of temporal validity. There was also a highly significant main effect of Duration, $F(2, 92) = 12.45$, $P < .001$, $\eta_p^2 = .21$, and an equally strong Duration × Time interaction, $F(38, 1748) = 13.02$, $P < .001$, $\eta_p^2 = .22$, which was dominated by differences between Short and Medium duration for the increase from time bins 7 to 8, $F(1, 46) = 14.08$ $P < .001$, $\eta_p^2 = .12$, differences between Medium and Long for the increase from time bins 4 to 5, $F(1, 46) = 15.10$, $P < .001$, $\eta_p^2 = .12$, the decreases from time bins 8 and 9, and 9 to 10, $F(1, 46) = 15.05$, $P < .001$, $\eta_p^2 = .12$, $F(1, 46) = 23.40$, $P < .001$, $\eta_p^2 = .12$, respectively. These effects were caused by staggered peaks and increasing durations of activation for the Short, Medium and Long radio speech conditions, respectively, clearly meeting the requirement of experimental validity. This provides strong support for a functional brain network reliably responding to radio speech, displaying strong spatial, temporal and experimental validity[38] (see table 1). No main effects or interactions involving Group were significant (all $P > .35$; see fig. 1C).

Component 3: focus on visual features. The anatomical regions associated with Component 3 are outlined in figure 2A, and the anatomical description is presented in supplementary table S7. The negative loadings provide a weak match to the Focus on Visual Features (FVF) network (Fisher's $z = .64$), providing weak evidence for spatial validity. Component 3 is characterized by bilateral deactivation in occipital areas such as the occipital pole (xyz: −27, −91, 16; 15, −88, 28). The FVF network is known to deactivate when the visual details of the task are not relevant to response.[36] [FVF,42] Figure S2,[46] Figure 5.60.

Figure 2B displays the estimated HDR shape for Component 3. Component 3 had a significant effect of Time, $F(19, 874) = 14.70$, $P < .001$, $\eta_p^2 = .24$, and although the HDR was biologically plausible, it did not provide a clear peak. The main effect of Duration was not significant ($P = .64$), but there was a significant Duration × Time interaction with a small effect, $F(38, 1748) = 2.31$, $P < .001$, $\eta_p^2 = .05$. This interaction was dominated by (1) a steeper increase/decrease for Short relative to Medium for the increase from time bins 4 to 5/5 to 6, respectively, $F(1, 46) = 6.12$, $P < .05$, $\eta_p^2 = .12$; $F(1, 46) = 4.45$, $P < .05$, $\eta_p^2 = .09$, respectively, due to an earlier peak for Short relative to Medium (time point 5 *vs*. 6, respectively), and (2) a steeper increase between Medium and Long from time bins 4 to 5, $F(1, 46) = 4.99$, $P < .05$, $\eta_p^2 = .10$, due to a peak at time point 6 for medium versus 8 for long. Therefore, these effects were caused by staggered peaks/increasing extensions of activation for the Short, Medium, and Long conditions, respectively, meeting experimental validity, but with a small effect size. This provides weaker evidence for a functional brain network reliably *deactivating* visual perception regions in response to auditorily presented stimuli. No main effects or interactions involving Group were significant (all $P > .05$; see figure 2C).

Melbourne and Utrecht hallucinations Merged (Short/Long)

For the merged analysis of the Melbourne and Utrecht data, whereby voice hearers indicated the start and end of hallucinations by button press or ball squeeze/release, respectively, 3 components were extracted from the task-related variance in BOLD signal, as determined by examining the scree plot.[44,45] Component 1 did not show temporal or spatial validity for the Melbourne or Utrecht

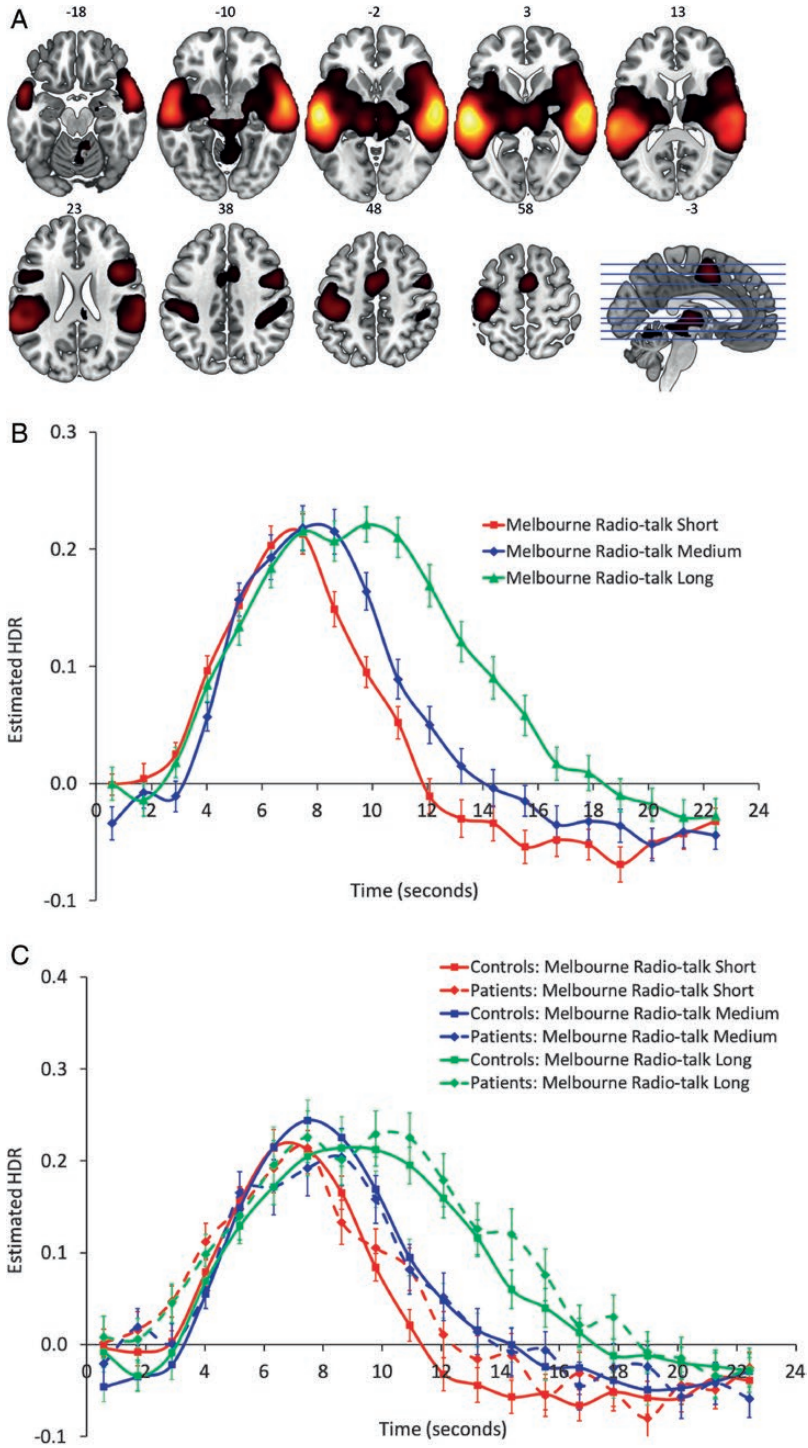

Fig. 1 (A) dominant 20% of component loadings for Component 2, proposed Auditory Perception network, from the Melbourne patient/control radio speech analysis. MNI *Z*-axis coordinates are displayed; left is left. Positive threshold = 0.11, max = 0.41. (B) mean finite impulse response (FIR)-based predictor weights plotted as a function of poststimulus time and condition. C (bottom): mean FIR-based predictor weights plotted as a function of post-stimulus time and condition shown with group differences. Error bars are standard errors.

data, as the HDR shape was not plausible for either site, and the brain images did not show recognizable anatomical patterns, so Component 1 is not reported here (see supplementary figure S5). Component 3 did not reveal a biologically plausible HDR shape for the Utrecht data ($P = .05$), but did for the Melbourne data, so it is reported here for the Melbourne data only, with the Utrecht data Component 3 reported in the supplementary material

(figure S6). Component 2 matched templates for the sensorimotor (response) network[33] Figure 7 and Table 6,[36] and exhibited plausible HDR shapes for both sites, meeting criteria for both spatial and temporal validity for both sites, so is reported below.

Component 2: sensorimotor (response) network. The anatomical regions associated with Component 2 are outlined in figure 3A, and the anatomical description of Component 2 is presented in supplementary table S8. Activation on this network involved bilateral pre- and post-central gyri (BAs 3, 4, 6), juxtapositional lobule cortex, and insular cortex (BA 47), which are all typical for the motor response network regions based on comparison with previous exemplar images.[33] Figure 7 and Table 6,[36]

Figure 3B displays the estimated HDR shape for Component 2 from the Melbourne hallucinations data, for which a significant effect of Time was found, $F(36, 396) = 2.20$, $P < .001$, $\eta_p^2 = .17$. However, there was no significant effect involving duration ($P > .25$) suggest that this HDR shape shows temporal validity, but not experimental validity.

For the Utrecht hallucinations data (HDR shown in figure 3C), a significant effect was found for Time $F(36, 504) = 8.41$, $P < .001$, $\eta_p^2 = .38$, showing reliability of HDR shape over participants. However, the absence of

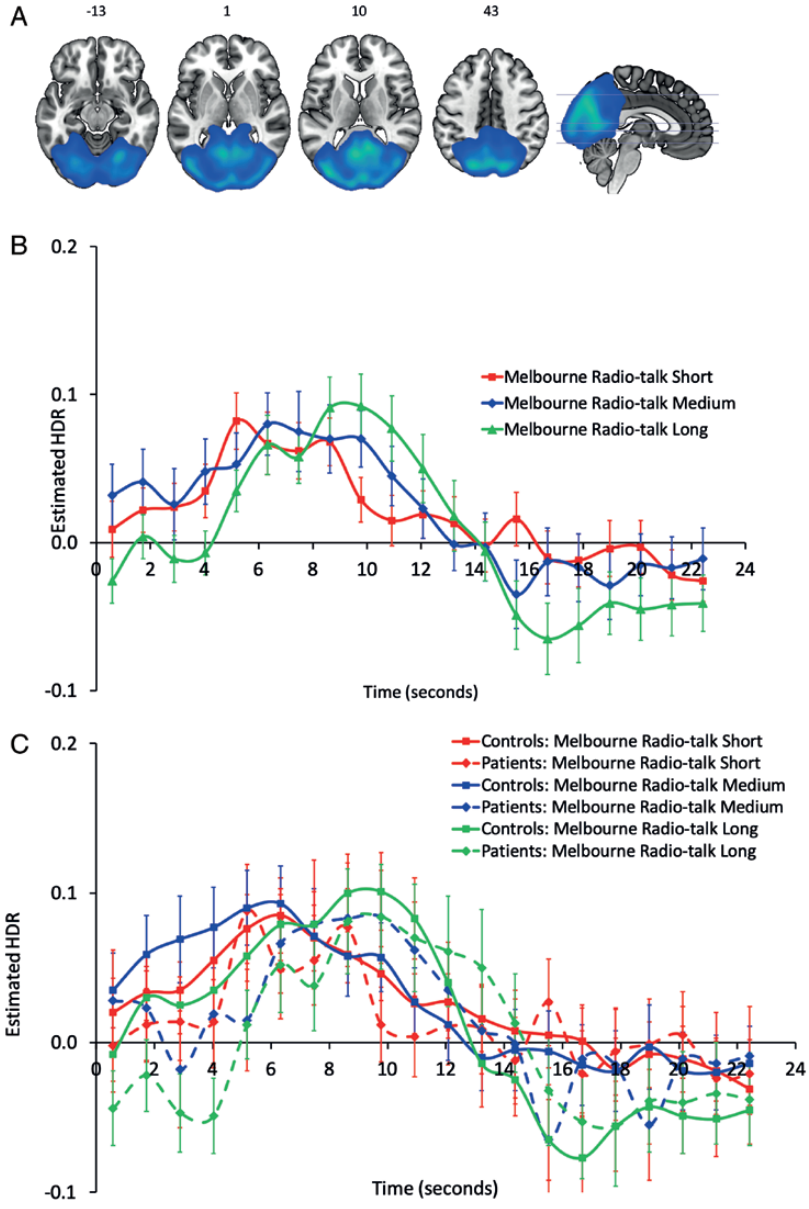

Fig. 2. (A) dominant 20% negative component loadings for Component 3, from the Melbourne radio speech analysis, Focus on Visual Features/Auditory Perception. MNI *Z*-axis coordinates are displayed; left is left. Negative threshold = −0.12, min = −0.19. (B) mean FIR-based predictor weights plotted as a function of post-stimulus time and condition. (C) mean FIR-based predictor weights plotted as a function of post-stimulus time and condition shown with group differences. Error bars are standard errors.

effects involving Duration ($P > .10$) suggests a failure of experimental validity (see table 1). Therefore, the pattern of activation in the response network of Utrecht participants also does not align with duration of reported hallucinations.

Component 3: focus on visual features. The anatomical regions associated with Component 3 are outlined in figure 4A, and the anatomical description of Component 3 is presented in supplementary table S10. The negative loadings matched the Focus on Visual Features (FVF) network, characterized by bilateral deactivation in occipital areas such as the occipital pole (xyz: 30, −91, −8; −27, −94, 1), which is known to deactivate when the visual details of the task are not relevant to response[36] FVF,[42] Figure S2,[46] Figure 5.60.

Figure 4B displays the estimated HDR shape for Component 3 for the Melbourne sample. For the Melbourne data, there was a biologically plausible HDR shape, and a significant effect of Time $F(36, 396) = 4.10$, $P < .05$, $\eta_p^2 = .27$, in the absence of significant effects involving Duration (all $P > .1$). There was not a biologically plausible HDR shape for the Utrecht data ($P = .05$; see supplementary figure S6).

Discussion

In the current study, fMRI data were collected during real-time symptom capture of AVHs at two sites (Melbourne and Utrecht), and merged, using fMRI-CPCA, a multidimensional analysis technique that can extract brain networks optimized to be predictable from the precise timing of the reported hallucinatory experiences. This study aimed to determine whether the functional brain networks that are detectable based on the timing of the reported experience of AVHs are underlying the hallucinatory experience itself or other cognitive events. The analyses were set up with clear criteria for spatial validity, temporal validity, and experimental validity, such that BOLD signal associated with hallucinations, if present, could be proven while avoiding the reverse inference fallacy by requiring duration-dependent HDRs. Although strong duration-dependent HDRs for radio speech perception was observed, this was not observed for hallucinations experiences. However, a network retrieved for AVH events anatomically matched the sensorimotor (response) network, for which the literature clearly demonstrates that this network is associated with the execution of response processes. This set of results suggests that fMRI may not be able to detect brain activity associated with the hallucination itself, but can readily detect the brain activity associated with generating responses indicating the start/end of an experienced hallucination.

Table 1 Experimental Validity (Duration × Time interaction) column shows that only radio speech elicited BOLD signal which was duration dependent. The hallucinations blocks did reliably elicit a HDR shape (Temporal Validity column), and when this involved activation (not deactivation), this always conformed closely to the known response network (Merged: Melbourne Hallucinations C2, and Merged: Utrecht Hallucinations C2). A clear response network did not emerge for Melbourne radio speech alone, but response network regions such as left dominated precentral and postcentral gyri (xyz: −36, −19, 64; −45, −28, 49; respectively)[33] Figure 7 and Table 6,[36] were included on the Auditory Perception component.

These results are difficult to reconcile with the many previous neuroimaging studies and meta-analyses have identified brain regions showing activation during AVHs as auditory perception related.[9,20–22,47] In supplementary table S11 we group together the brain regions found to be involved during AVHs from three meta-analyses,[20–22] in comparison to brain regions concluded to be implicated as part of the response network from the current study, and another study analyzed using fMRI-CPCA with an empirically derived response network.[33] From supplementary table S11, it can be seen that many brain regions implicated for AVHs overlap with the response network observed in component 2 from the Melbourne and Utrecht patient hallucinations (S/L) analysis. For example, the left insula (*x y z* peak near: −42 4 −2) has been shown to be implicated in AVH by the meta-analyses[20–22]; however, in the current analysis and Sanford et al. (2020), this region is considered to be a part of the response network. Similarly, activation near peaks in the right and left post central gyrus is seen in some of the meta-analyses and the response network identified form this analysis. In regard to the STG, all three meta-analyses have shown peaks of activation in this region.[20–22] Although peak activation was not observed in this region in the response network from Sanford et al. (2020), peak activation was observed in adjacent brain areas to the STG, in the left central opercular cortex (*x y z* peak: −54 −19 16) and the right inferior frontal gyrus (*x y z* peak: 57 14 −2) from component 2 of the Melbourne and Utrecht patient hallucinations (S/L) analysis. However, local activation near to the STG is not, on its own, evidence for an underlying voice perception network. Duration-dependent signal in the HDR for hallucinations (absence of experimental validity, see table 1 rightmost column) is also required. A recent study also reported somatosensory regions as responsible for hallucinations without factoring in variability in the duration of voices, or the number of button presses between compared conditions.[48]

Comparing to the previously published version of the Utrecht data,[43] they also report motor areas associated with timing of AVH events. The authors of this work wrote: "While activation of motor areas, as observed in this study, most likely results from the employed balloon squeeze paradigm, the role of bilateral frontal and temporoparietal regions in the experience of AVH is not

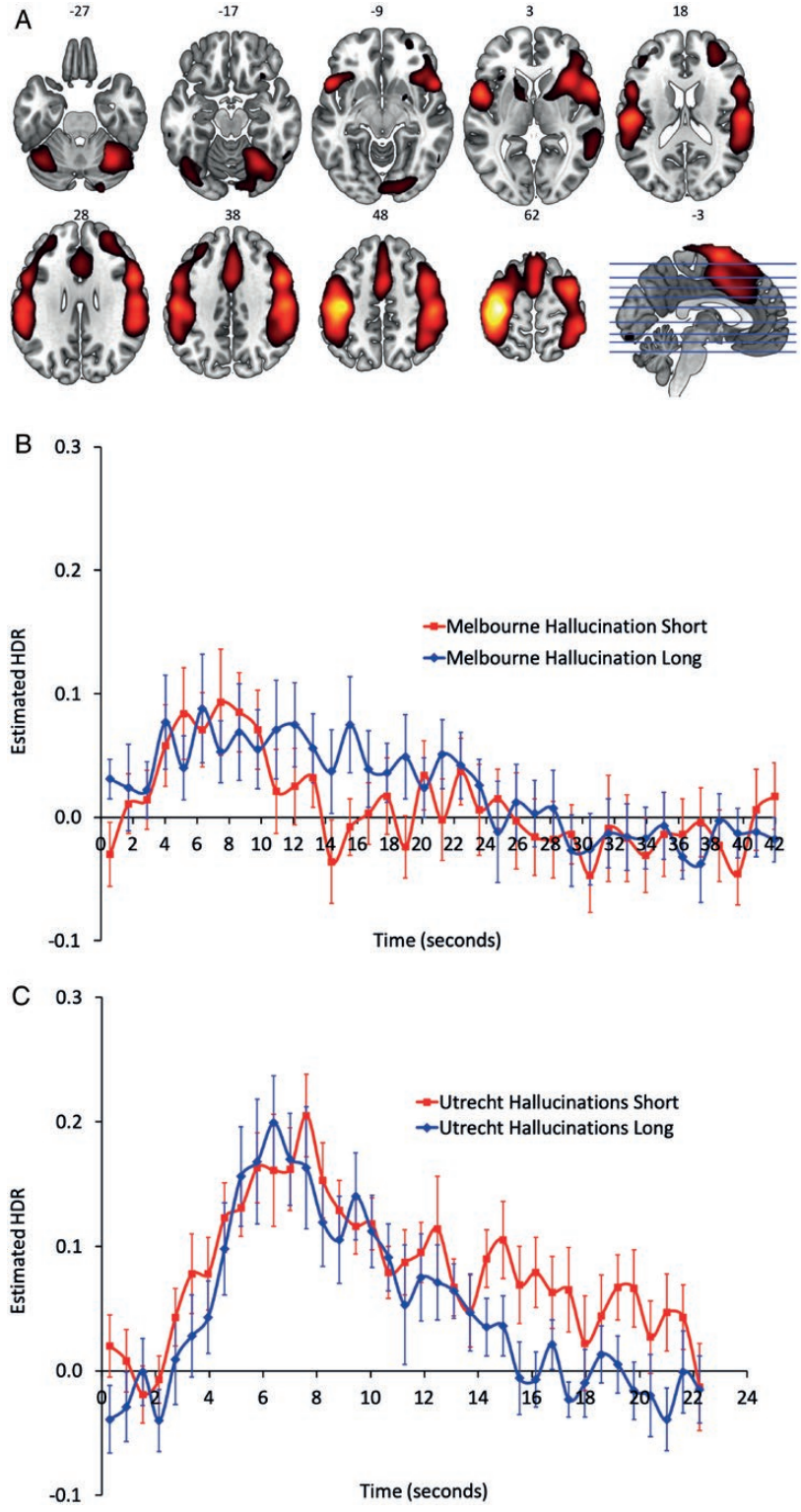

Fig. 3. (A) dominant 20% of component loadings for Component 2, proposed one-handed response network, from the Melbourne and Utrecht patient hallucinations merged analysis. MNI *Z*-axis coordinates are displayed; left is left. Positive threshold = 0.10, max = 0.30. (B) mean FIR-based predictor weights plotted as a function of post-stimulus time and condition for Melbourne data. (C) mean FIR-based predictor weights plotted as a function of post-stimulus time and condition for Utrecht data. Error bars are standard errors.

yet clear." (p. 1079). These named anatomical regions were assumed *not* to be motor, so it was suggested that they may be related to hallucinations. We also found these activations in the current paper (e.g., see slice 3 in figure 3); however, (1) they did not show duration dependence, and (2) these regions do fall on the motor network,

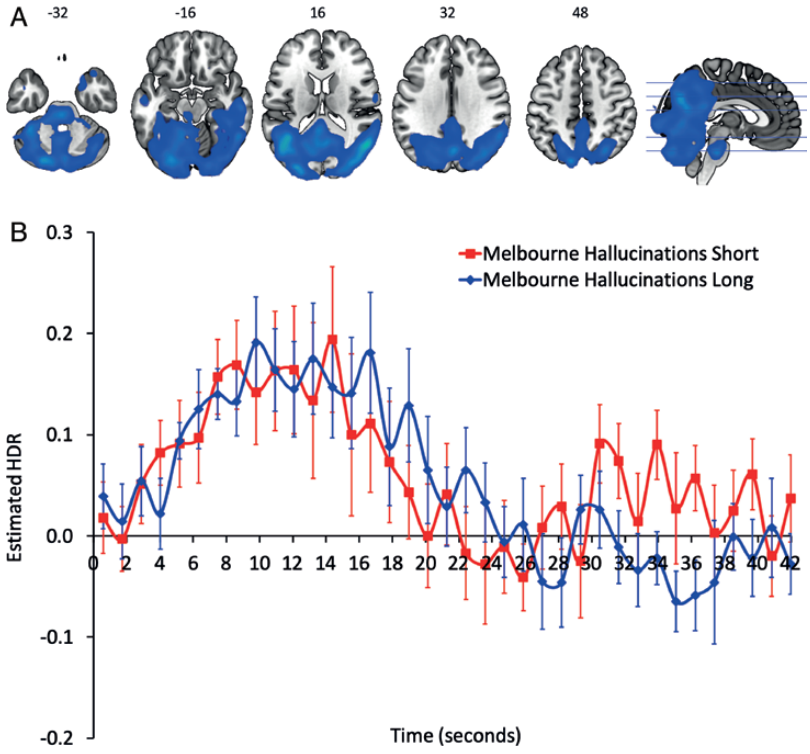

Fig. 4. (A) dominant 20% of component loadings for Component 3, proposed Focus on Visual Features network, from the Melbourne patient hallucinations merged analysis. MNI *Z*-axis coordinates are displayed; left is left. Negative threshold = −0.10, min = −0.18. (B) mean FIR-based predictor weights plotted as a function of post-stimulus time and condition (error bars are standard errors) for Melbourne data.

as well as on a number of other networks (Percival et al., 2020), which is why we hold that anatomical identification alone is insufficient for identification of brain function (reverse inference fallacy).

Limitations

This study was subject to some key limitations, many of which are common to all symptom capture studies. First, the average duration of hallucinations experienced by participants varied greatly, with some participants reporting many hallucinations less than 3 s long, and others having relatively long hallucinations. In addition, the method for self-reporting hallucinations may vary between individuals, with some reporting more frequently (e.g., between individual words), with others reporting rare events (e.g., after longer sentences), therefore making it difficult to average over trials within participants. This inconsistency between participants may further make it more difficult to recognize functional brain networks involved in hallucinations.

Detection of AVHs with fMRI requires repeated consistent changes in HDR over multiple occasions. However, it may be that hallucinations are more readily detectable when averaged over a longer period of time, and could never elicit a duration-dependent HDR in fMRI. For example, positron emission tomography averages brain activity over minutes rather than seconds, which negates the requirement for participants to provide precise temporal information about AVHs on the magnitude of seconds.[49,50] It is also possible to use the postscan epoch self-report of hallucinations to average over a period of time than individual events in fMRI.[24,51]

Recommendations

It is possible that fMRI cannot detect AVHs, which would point to EEG or MEG as better candidates for hallucinations capture studies. However, future hallucination capture studies using fMRI would benefit from adjusted experimental designs to reveal better information about brain networks involved during the experience of hallucinations. There are a number of different ways that hallucinations capture fMRI studies would be done differently in the future to facilitate interpretation:

(1) Since it will be necessary to classify actual hallucinations into short/long, training is essential to ensure the duration of the voice matches onto the duration of the press/squeeze on/off. Clear instructions should be given to participants on how to indicate hallucination onset and offset. Participants should take part in intensive practice/training sessions involving sound files, silent thinking, and pressing for hallucinating voices to make sure it is very clear how to indicate the beginning and end of an event. It is important to avoid frequent and repeated squeezes/presses. For example, instruct the

subjects to maintain the squeeze/press until there have been no voices for 2 s, that way, the shortest squeeze/press duration will be 2 s. The offset could be indicated by release of a press rather than another different press. If someone has no 2 s break from voices over the entire run, they will be squeezing the entire run, and the run will not be usable anyway, but that can't be avoided. Data should be collected on these practice/training sessions showing that participants understood the instructions.

(2) To avoid excessive pressing/squeezing, onset and offset of hallucinations (and radio speech or inner speech) should be indicated by onset and offset of button presses/ball squeezes, with a minimum press of 2 s.

(3) Alternating blocks (perhaps 2 min long) of pressing for hearing radio voice (e.g., 2 s vs. 6 s) inner speech (e.g., counting 2 s vs. 6 s) for hallucinations (minimum 2 s), and pressing for counted 2 s vs. 6 s following a cue, can be compared to the same condition but without the response. 30 second rest periods can be used to allow estimation of the response network to separate from any inner speech/speech perception and hallucinations networks. There may be hallucinations when hearing radio, but the timing will match the radio not the hallucinations. Different lengths of radio sentences/inner speech/hallucinations allow networks to be elicited with staggered peaks on the HDR to ensure experimental validity. Alternatively, attempts could be made to match the number, duration, and timing of inner speech events to the timing of the hallucination events.[52]

(4) Cue the start and stop press (or squeeze), or no response, with an auditory cue, which promotes the squeeze release in the absence of internal or external speech.

(5) ITIs are very important in fMRI, and during inner speech or speech perception or button pressing, there should be a distribution of mostly short (2, 4 s), but a few long (6, 8 s) ITIs.[53]

Conclusion

In this fMRI study, we addressed: (1) the experimental/timing confound of AVH events with response events by using a dimensional analysis method, (2) likely mutual conflation of AVH and response events by using a FIR model, and (3) the fallacy of reverse inference by requiring experimental validity as well as anatomical validity as sufficient evidence for detection of AVH events. The auditory perception network revealed a speech-duration-dependent HDR signal when radio clips were heard, but under no conditions were duration-dependent HDRs elicited during online-reported hallucinations. In contrast, an anatomical depiction of the response network was observed for button press or squeeze response when analyzing the hallucinations from merging the Utrecht and Melbourne datasets together. No brain networks were clearly demonstrated to be sensitive to the experience of hallucinations themselves, because duration-dependent fMRI signal was not observed for any of the components. Since responses are perfectly confounded with hallucination onsets, there is no strong evidence that event-related symptom-capture fMRI-paradigms can detect brain networks involved in hallucinations over and above response processes. This does not imply that neurostimulation methods targeting the STG are invalid,[54–57] or that hallucinations do not involve the STG, but simply suggests that either event-related fMRI cannot detect hallucinations, or it cannot detect the duration of hallucinations, or different designs may be required to indicate the onset and offset of hallucinations.

Supplementary Material

Supplementary data are available at *Schizophrenia Bulletin Open* online.

Funding

This work was supported the Australian National Health and Medical Research Council (NHMRC; senior research fellowship to S.L.R (ID: 1154651), a project grant to S.L.R, M.H. and W.W (ID: 1060664)), the Programme Hospitalier de Recherche Clinique National (PHRC-N) Grant MULTIMODHAL 2013 to R.J., and the Dutch Research Council NOW to I.S.

Acknowledgments

We thank the Australian National Imaging Facility for their support, and Andre Zamani for commenting on this manuscript. The authors declare that they have no conflicts of interest.

References

1. Bauer SM, Schanda H, Karakula H, *et al.* Culture and the prevalence of hallucinations in schizophrenia. *Compr Psychiatry.* 2011;52(3):319–325.
2. Lim A, Hoek HW, Deen ML, *et al.* Prevalence and classification of hallucinations in multiple sensory modalities in schizophrenia spectrum disorders. *Schizophr Res.* 2016;176(2):493–499.
3. Rapin L, Loevenbruck H, Dohen M, Metzak PD, Whitman JC, Woodward TS. Hyperintensity of functional networks involving voice-selective cortical regions during silent thought in schizophrenia. *Psychiatry Res Neuroimaging* 2012;202(2):110–117.
4. Ford JM, Roach BJ, Faustman WO, Mathalon DH. Synch before you speak: auditory hallucinations in schizophrenia. *Am J Psychiatry.* 2007;164(3):458–466.
5. Northoff G, Qin P. How can the brain's resting state activity generate hallucinations? A resting state hypothesis of auditory verbal hallucinations. *Schizophr Res.* 2011;127(1–3):202–214.

6. Lavigne K, Woodward TS. Hallucination- and speech-specific hypercoupling in frontotemporal auditory and language networks in schizophrenia using combined task-based fMRI data: an fBIRN study. *Hum Brain Mapp.* 2018;39:1582–1595.
7. Lavigne KM, Rapin LA, Metzak PM, *et al.* Left-dominant temporal-frontal hypercoupling in schizophrenia patients with hallucinations during speech perception. *Schizophr Bull.* 2015;41(1):259–267.
8. Suzuki M, Yuasa S, Minabe Y, Murata M, Kurachi M. Left superior temporal blood flow increases in schizophrenic and schizophreniform patients with auditory hallucination: a longitudinal case study using 123I-IMP SPECT. *Eur Arch Psychiatry Clin Neurosci.* 1993;242(5):257–261.
9. Sommer IE, Diederen KM, Blom JD, *et al.* Auditory verbal hallucinations predominantly activate the right inferior frontal area. *Brain* 2008;131(Pt 12):3169–3177.
10. Hoffman RE, Hawkins KA, Gueorguieva R, *et al.* Transcranial magnetic stimulation of left temporoparietal cortex and medication-resistant auditory hallucinations. *Arch Gen Psychiatry.* 2003;60(1):49–56.
11. Vercammen A, Knegtering H, Liemburg EJ, den Boer JA, Aleman A. Functional connectivity of the temporo-parietal region in schizophrenia: effects of rTMS treatment of auditory hallucinations. *J Psychiatr Res.* 2010;44(11):725–731.
12. Jardri R, Delevoye-Turrell Y, Lucas B, *et al.* Clinical practice of rTMS reveals a functional dissociation between agency and hallucinations in schizophrenia. *Neuropsychologia* 2009;47(1):132–138.
13. Ford JM, Hoffman RE. Functional brain imaging of auditory hallucinations: from self-monitoring deficits to co-opted neural resources. In: Jardri R, Pins D, Cachia A, Thomas P, eds. *The Neuroscience of Hallucinations*. London: Springer; 2013.
14. Hoffman R. Comment on Vercammen, Knegtering, den Boer, Liemburg, & Aleman submitted 29 April, 2010. *Schizophrenia Research Forum* 2010(Apr 29, 2010).
15. David AS. The neuropsychological origin of auditory hallucinations. In: David AS, Cutting JC, eds. *The Neuropsychology of Schizophrenia*. Hillside, NJ: Erlbaum; 1994:269–313.
16. Curcic-Blake B, Ford JM, Hubl D, *et al.* Interaction of language, auditory and memory brain networks in auditory verbal hallucinations. *Prog Neurobiol.* 2017;148:1–20.
17. Allen P, Larøi F, McGuire PK, Aleman A. The hallucinating brain: a review of structural and functional neuroimaging studies of hallucinations. *Neurosci Biobehav Rev.* 2008;32(1):175–191.
18. Moritz S, Laroi F. Differences and similarities in the sensory and cognitive signatures of voice-hearing, intrusions and thoughts. *Schizophr Res.* 2008;102(1-3):96–107.
19. Bentall RP. The illusion of reality: a review and integration of psychological research on hallucinations. *Psychol Bull.* 1990;107(1):82–95.
20. Kompus K, Westerhausen R, Hugdahl K. The "paradoxical" engagement of the primary auditory cortex in patients with auditory verbal hallucinations: a meta-analysis of functional neuroimaging studies. *Neuropsychologia* 2011;49(12):3361–3369.
21. Zmigrod L, Garrison JR, Carr J, Simons JS. The neural mechanisms of hallucinations: a quantitative meta-analysis of neuroimaging studies. *Neurosci Biobehav Rev.* 2016;69:113–123.
22. Jardri R, Pouchet A, Pins D, Thomas P. Cortical activations during auditory verbal hallucinations in schizophrenia: a coordinate-based meta-analysis. *Am J Psychiatry.* 2011;168(1):73–81.
23. Leroy A, Foucher JR, Pins D, *et al.* fMRI capture of auditory hallucinations: validation of the two-steps method. *Hum Brain Mapp.* 2017;38(10):4966–4979.
24. Fovet T, Yger P, Lopes R, *et al.* Decoding activity in Broca's area predicts the occurrence of auditory hallucinations across subjects. *Biol Psychiatry.* 2022;91(2):194–201.
25. Jardri R, Thomas P, Delmaire C, Delion P, Pins D. The neurodynamic organization of modality-dependent hallucinations. *Cerebral Cortex (New York, NY: 1991)* 2013;23(5):1108–1117.
26. Friston K. Statistical parametric mapping. In: Friston K, Ashburner J, Kiebel S, Nichols T, Penny W, eds. *Statistical Parametric Mapping - The Analysis of Functional Brain Images*; 2007:10–31.
27. Poldrack RA. Can cognitive processes be inferred from neuroimaging data? *Trends Cogn Sci.* 2006;10(2):59–63.
28. Poldrack RA. Mapping mental function to brain structure: how can cognitive neuroimaging succeed? *Perspect Psychol Sci.* 2010;5:753–761.
29. Henson R. Forward inference using functional neuroimaging: dissociations versus associations. *Trends Cogn Sci.* 2006;10(2):64–69.
30. Harley TA. Promises, promises. *Cogn Neuropsychol.* 2004;21(1):51–56.
31. Aguirre GK. Functional imaging in behavioral neurology and neuropsychology. In: Feinberg TE, Farah MJ, eds. *Neurology and Neuropsychology*. 2nd ed. McGraw-Hill; 2003:85–96.
32. Sanford N, Woodward TS. Functional delineation of prefrontal networks underlying working memory in schizophrenia: a cross-dataset examination. *J Cogn Neurosci.* 2021;33:1880–1908.
33. Sanford N, Whitman JC, Woodward TS. Task merging for finer separation of functional brain networks in working memory. *Cortex.* 2020;125:246–271.
34. Woodward TS, Feredoes E, Metzak PD, Takane Y, Manoach DS. Epoch-specific functional networks involved in working memory. *Neuroimage* 2013;65:529–539.
35. Metzak PD, Feredoes E, Takane Y, *et al.* Constrained principal component analysis reveals functionally connected load-dependent networks involved in multiple stages of working memory. *Hum Brain Mapp.* 2011;32(6):856–871.
36. Percival CM, Zahid HB, Woodward TS. Set of task-based functional brain networks derived from averaging results of multiple fMRI-CPCA studies: CNoS-Lab/Woodward_Atlas. Zenodo.; 2020.
37. Woodward TS, Zahid H, Percival CM, Woodward TS, Zahid H, Percival CM, Woodward TS, Zahid H, Percival CMs. Brain Network Classification, 2021.
38. Ribary U, Mackay AL, Rauscher A, *et al.* Emerging neuroimaging technologies: towards future personalized diagnostics, prognosis, targeted intervention and ethical challenges. In: Illes J, ed. *Neuroethics II: Defining the Issues in Theory, Practice, and Policy*. Oxford: Oxford University Press; 2017:15–52.
39. Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc Natl Acad Sci USA.* 2005;102(27):9673–9678.
40. Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL. A default mode of brain function. *Proc Natl Acad Sci USA.* 2001;98(2):676–682.
41. Yeo BT, Krienen FM, Sepulcre J, *et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysiol.* 2011;106(3):1125–1165.

42. Sanford N, Whitman JC, Woodward TS. (Supplementary Data) Task merging for finer separation of functional brain networks in working memory. *Cortex* 2020;125:246-271.
43. Diederen KMJ, Daalman K, de Weijer AD, *et al.* Auditory hallucinations elicit similar brain activation in psychotic and nonpsychotic individuals. *Schizophr Bull.* 2012;38(5):1074–1082.
44. Cattell R. The scree test for the number of factors. *Multivariate Behavioural Research*. 1966;1:245–276.
45. Cattell RA. Comprehensive trial of the scree and Kg criteria for determining the number of factors. *Multivariate Behav Res.* 1977;12(3):289–325.
46. Sanford N. *Functional Brain Networks Underlying Working Memory Performance in Schizophrenia: A Multi-experiment Approach*. Vancouver, Canada: Department of Psychiatry, University of British Columbia; 2019.
47. Suzuki M, Yuasa S, Minabe Y, Murata M, Kurachi M. Left superior temporal blood flow increases in schizophrenic and schizophreniform patients with auditory hallucination: a longitudinal case study using 123I-IMP SPECT. *Eur Arch Psychiatry Neurol Sci.* 1993;242(5):257–261.
48. Strawson WH, Wang HT, Quadt L, *et al.* Voice hearing in Borderline Personality Disorder across perceptual, subjective and neural dimensions. *Int J Neuropsychopharmacol.* 2021;25(5):375-386.
49. Silbersweig DA, Stern E, Frith C, *et al.* A functional neuroanatomy of hallucinations in schizophrenia. *Nature* 1995;378:176–179.
50. Silbersweig D. From symptom-capture neuroimaging to imaging biomarker development: the challenge of auditory hallucinations in schizophrenia. *Biol Psychiatry (1969)* 2022;91(2):164–166.
51. Leroy A, Foucher JR, Pins D, *et al.* fMRI capture of auditory hallucinations: validation of the two-steps method. *Hum Brain Mapp.* 2017;38(10):4966–4979.
52. Ellamil M, Fox KC, Dixon ML, *et al.* Dynamics of neural recruitment surrounding the spontaneous arising of thoughts in experienced mindfulness practitioners. *Neuroimage* 2016;136:186–196.
53. Zarahn E, Aguirre G, D'Esposito M. A trial-based experimental design for fMRI. *Neuroimage* 1997;6(2):122–138.
54. Brunelin J, Mondino M, Gassab L, *et al.* Examining transcranial direct-current stimulation (tDCS) as a treatment for hallucinations in schizophrenia. *Am J Psychiatry.* 2012;169(7):719–724.
55. Kar SK, Singh A, Prakash AJ. Neuromodulation in schizophrenia: relevance of neuroimaging. *Curr Behav Neurosci Rep.* 2020;7(3):139–146.
56. Mondino M, Jardri R, Suaud-Chagny M-F, Saoud M, Poulet E, Brunelin J. Effects of fronto-temporal transcranial direct current stimulation on auditory verbal hallucinations and resting-state functional connectivity of the left temporo-parietal junction in patients with schizophrenia. *Schizophr Bull.* 2016;42(2):318–326.
57. Yang F, Fang X, Tang W, *et al.* Effects and potential mechanisms of transcranial direct current stimulation (tDCS) on auditory hallucinations: a meta-analysis. *Psychiatry Res.* 2019;273:343–349.