Neuroscience and Biobehavioral Reviews 131 (2021) 663–687

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

journal homepage: www.elsevier.com/locate/neubiorev

External speech processing and auditory verbal hallucinations: A systematic review of functional neuroimaging studies

Sophie E. Richards [a,*], Matthew E. Hughes [a], Todd S. Woodward [b,c], Susan L. Rossell [a,d,1], Sean P. Carruthers [a,1]

[a] Centre for Mental Health, Faculty of Health, Arts & Design, Swinburne University of Technology, VIC, 3122, Australia
[b] Department of Psychiatry, University of British Colombia, Vancouver, BC, Canada
[c] BC Mental Health and Addictions Research Institute, Vancouver, BC, Canada
[d] Department of Psychiatry, St Vincent's Hospital, Melbourne, VIC, Australia

ARTICLE INFO

Keywords:
Auditory verbal hallucination
External speech
Listening
Non-clinical voice-hearers
Functional neuroimaging
Schizophrenia
Schizoaffective disorder
Voice-hearing

ABSTRACT

It has been documented that individuals who hear auditory verbal hallucinations (AVH) exhibit diminished capabilities in processing external speech. While functional neuroimaging studies have attempted to characterise the cortical regions and networks facilitating these deficits in a bid to understand AVH, considerable methodological heterogeneity has prevented a consensus being reached. The current systematic review investigated the neurobiological underpinnings of external speech processing deficits in voice-hearers in 38 studies published between January 1990 to June 2020. AVH-specific deviations in the activity and lateralisation of the temporal auditory regions were apparent when processing speech sounds, words and sentences. During active or affective listening tasks, functional connectivity changes arose within the language, limbic and default mode networks. However, poor study quality and lack of replicable results plague the field. A detailed list of recommendations has been provided to improve the quality of future research on this topic.

1. Introduction

Auditory verbal hallucination (AVH), otherwise referred to as voice-hearing, involves the perception of voices in the absence of an external auditory input (Slade and Bentall, 1988). At a global lifetime prevalence of roughly 5–15 % (Beavan et al., 2011), voice-hearing is reported in numerous clinical groups (de Leede-Smith and Barkus, 2013; Merrett et al., 2016; Toh et al., 2020b) alongside many individuals without a need for care (*i.e.* people who hear voices who are otherwise healthy; non-clinical voice-hearers; Johns et al., 2014; Peters et al., 2017; Van Os and Reininghaus, 2016). Being as varied and idiosyncratic as spoken language (Aleman and Larøi, 2008), voice-hearing is difficult to characterise. For example, a word, phrase or discourse may be perceived as originating from a sole speaker or multiple speakers talking independently, conversationally or simultaneously (McCarthy-Jones and Resnick, 2014). The accent, personification and spoken content of voices vary widely (Aleman and Larøi, 2008), as does their effect on the experiencer (Upthegrove et al., 2016). However, common phenomenological themes do exist (Waters and Fernyhough, 2017); for instance, most voice-hearers perceive voices as speaking at a normal volume (Larøi et al., 2012) and in either the second or third person (Tovar et al., 2019). Furthermore, most AVH are reminiscent of social interactions (Behrendt and Young, 2004) and have complex identities, whether these are constructed or likened to another person (Nayani and David, 1996; Romme et al., 2009).

Alongside the complex phenomenological experience of AVH, it has been argued that individuals who experience voice-hearing commonly exhibit a diminished ability to process external speech (Hugdahl and Sommer, 2018; Upthegrove et al., 2016). Such deficits include poor speech-related verbal attention and working memory (Daalman et al., 2011; Toh et al., 2020a), voice recognition (Johns et al., 2001; Woodward et al., 2007), and prosodic interpretation (Costafreda et al., 2008; Rossell and Boundy, 2005). The magnitude of speech processing impairments are often robustly linked to indices of voice-hearing severity such as their frequency, duration or content (Alderson-Day et al., 2017; Hugdahl and Sommer, 2018; Rossell et al., 2013). Such well-established links between external speech processing deficits and AVH have led researchers to examine the underlying neurobiological relationships

* Corresponding author.
E-mail address: serichards@swin.edu.au (S.E. Richards).
[1] Joint senior authors.

https://doi.org/10.1016/j.neubiorev.2021.09.006
Received 16 February 2021; Received in revised form 31 August 2021; Accepted 3 September 2021
Available online 10 September 2021

between the two processes (Ćurčić-Blake et al., 2017; Jardri et al., 2011a; Uhlhaas and Singer, 2010).

A wealth of literature has demonstrated the involvement of functional brain networks associated with speech processing, such as the auditory and language processing networks, in the pathophysiology of voice-hearing (Kang et al., 2009; Lavigne et al., 2015; Zhang et al., 2008a). However, considerable methodological heterogeneity between studies has impeded a consensus of findings being reached, perhaps due to differences in the activation paradigms used. For example, some studies have used emotionally neutral speech stimuli (Woodruff et al., 1997), others have used highly affective speech stimuli (Sanjuan et al., 2007), compared ear of presentation (Kompus et al., 2013) or probed speech-related memory (Ganguli et al., 1997). This has led to some contention regarding the degree of involvement of several cortical structures. For instance, fMRI signal intensity in the right anterior cingulate cortex (ACC) has been reported to both increase (Lewis-Hanna et al., 2011) and decrease (Horga et al., 2014b) in response to speech in voice-hearing samples. In EEG studies, auditory-evoked N100 amplitudes over temporal electrode sites have been reported to both increase and decrease while listening to speech sounds (Ford et al., 2002; Pinheiro et al., 2018; Thiebes et al., 2018). A systematic review of this literature would provide a distinct opportunity to disentangle the relationship between external speech processing and neural correlates in those that hear voices.

As such, the current review aims to synthesize the current understanding of the neurobiological underpinnings of external speech processing deficits which are specific to voice-hearing. To this end, the review will seek: (i) to assess the cortical regions of interest and functional networks persistently implicated in the processing of external speech and speech sounds among voice-hearers and; (ii) any cortical regions or networks that may be particularly affected by different speech components. To provide this synthesis, studies were grouped by approximating across the hierarchy of language processing: basic speech sounds, single words and finally, strings of words and sentences. Within each category, these tasks have been ordered by their cognitive (*e.g.* passive or active listening) or affective load to discriminate tasks which include these additional levels of processing. Non-AVH participants with a psychiatric diagnosis (*e.g.* schizophrenia) were described as either being state-negative (*i.e.* a past history of AVH, however not actively hearing voices) or trait-negative (*i.e.* no history of AVH). A secondary aim will focus on study-specific demographic and methodological variables that may underlie heterogeneity between study outcomes. Following this synthesis, future directions for imperative, multifaceted research approaches will be discussed.

2. Methods

2.1. Search protocol

Selection of reports was conducted as per the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Liberati et al., 2009) statement where appropriate; a protocol is registered with the International Prospective Register of Systematic Reviews (PROSPERO; registration number: CRD42020140409). Using the online databases of PubMed and Scopus and with additional searches through relevant review articles and publications, eligible literature was constrained to those published after the 1st of January 1990 and before the 10th of June 2020. Search terms focussed around AVH, functional neuroimaging and external voice tasks, with appropriate truncation and syntax used to optimise searches in each database (refer to Supplementary Material I for full search syntax).

2.2. Study selection & eligibility criteria

Criteria for inclusion were established across two screening stages. At stage one, one reviewer confirmed eligibility through titles, abstracts and keywords for all publications that: (a) were written in English and empirical; (b) investigated AVH and; (c) external voice processing and; (d) utilized one or more functional neuroimaging modality. At stage two, full-length publications were assessed independently by two reviewers and deemed eligible if: (e) an external voice task had been performed inside of one or more of the following: electroencephalography (EEG); functional magnetic resonance imaging (fMRI); magnetoencephalography (MEG); proton emission spectography (PET) or; single proton emission computed tomography (SPECT) and; (f) sufficient detail existed for distinction of an AVH group within the study population. Articles that used a single case-report or were reviews or meta-analyses were excluded. Discrepancies between final stage two screening were resolved by discussion, with a third author acting as adjudicator for uncertain records.

2.3. Data extraction and synthesis

Data extraction was performed by one reviewer using a checklist approved by all authors and inclusive of all items set out in PRISMA-P item 12. Extracted items included publication-specific data (*e.g.* study design, comparative approaches); demographics (*e.g.* sample size, AVH classification); external voice task variables (*e.g.* response time, accuracy) and; neuroimaging-specific measures obtained. It was assumed that all listening tasks were presented binaurally unless otherwise specified. Where possible, mean differences and standard deviations were extracted or calculated as primary summary measures. Upon request, several authors provided unpublished information (Ford et al., 2002; Jardri et al., 2011a; Lavigne and Woodward, 2018; Mechelli et al., 2007; Pinheiro et al., 2017; Sanjuan et al., 2007; Stephane et al., 2018). Due to the considerable methodological variability and subsequent limited overlap between studies, a meta-analysis was deemed impractical. In response, a structured narrative-based synthesis of results was utilized.

2.4. Quality assessment of individual studies

The current review developed a tool to evaluate the quality of each eligible study based on the size and characterisation of their AVH samples and descriptions of the external speech processing task. The quality of reporting of neuroimaging variables was also assessed, based on current recommendations for fMRI (Bradshaw et al., 2017; Poldrack et al., 2017, 2008), PET and SPECT (Egerton et al., 2017; Kaneta, 2020; Kapucu et al., 2009; Silverman, 2004) and EEG and MEG (Anderer et al., 1996; Keil et al., 2014; Puce and Hämäläinen, 2017) studies. The full details of the quality assessment tools are presented in Supplementary Material II.

3. Results

3.1. Study selection

The electronic database searches provided a total of 707 publications, with 608 remaining after removal of duplicates. After reviewing titles, abstracts and keywords for exclusion criteria, a total of 94 publications were deemed eligible and selected for full text screening. Of these, 38 studies were identified for inclusion in this systematic review. A flowchart of this process can be seen in Fig. 1.

3.2. Study characteristics

Studies predominantly investigated voice-hearers within schizophrenia spectrum disorder (SSD). Twenty-three studies assessed voice-hearers with a current diagnosis of schizophrenia (Sz), three studies assessed combined SSD groups, three assessed voice-hearers in their first or second psychotic episode and a further three assessed individuals with Sz who were not hearing voices at the time of testing. Five studies assessed voice-hearing in non-clinical voice-hearing samples. Of these, three assessed AVH during normal wakefulness (Alderson-Day et al., 2017; Kompus et al., 2013; Pinheiro et al., 2018), one assessed a

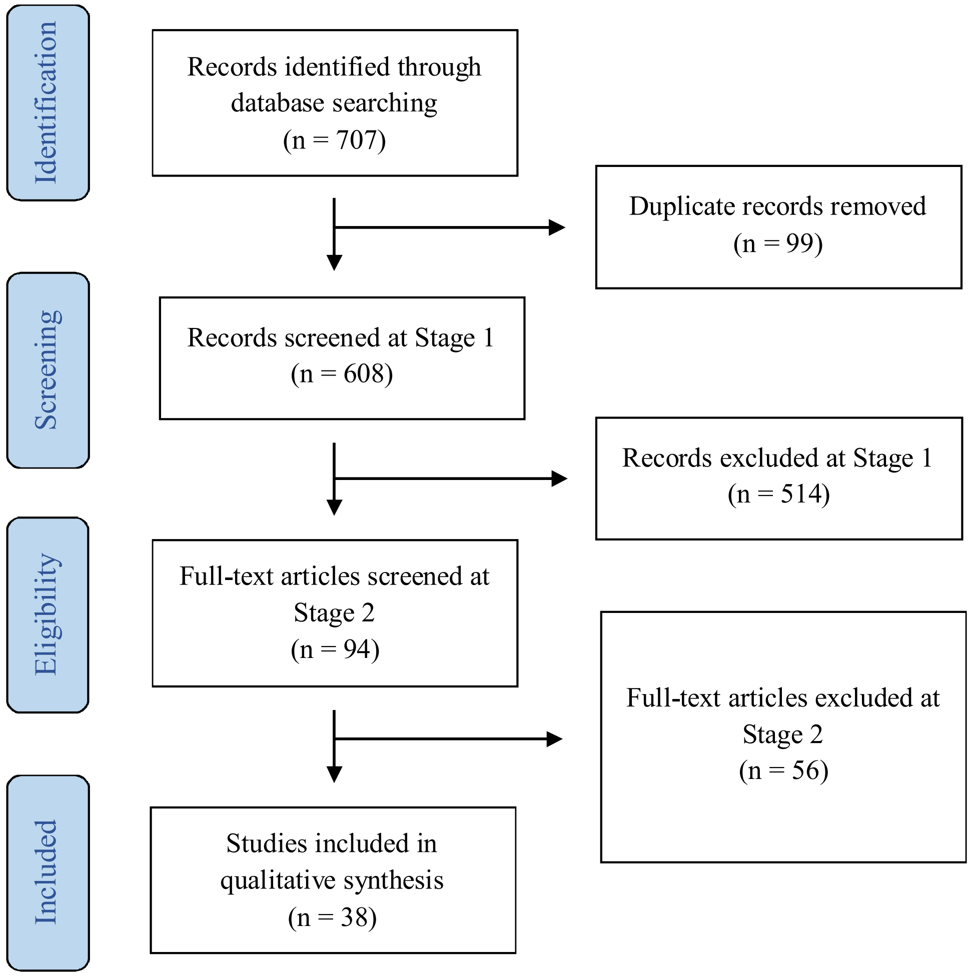

Fig. 1. PRISMA flow diagram of the article screening and selection process.

hypnogogic or hypnopompic AVH sample (Lewis-Hanna et al., 2011) and one induced AVH with intravenous ketamine administration (Thiebes et al., 2018). One final study assessed AVH within an epilepsy sample (Korsnes et al., 2010).

Brain function was assessed with fMRI in 25 studies, EEG in 10 studies, PET in one study, MEG in one study and SPECT in one study. Of the 25 fMRI studies, 12 used seed-based or region of interest (ROI) approaches, 12 used whole brain analyses, and one did not state their acquisition dimensions. Seven of these studies specifically examined functional networks and one correlated fMRI data against structural MRI data. Among encephalographic studies, eight analysed auditory-evoked event-related potentials (ERPs), with five also performing time-frequency analyses. The one PET study implemented 2-deoxy-2-[fluorine-18]fluoro-D-glucose (^{18}F-FDG) to investigate relative glucose metabolism rates and finally, the one SPECT study used 15-Oxygen labelled water ([^{15}O]-H_2O) to investigate functioning with rCBF.

3.3. Summary of methodological quality

The full details of the quality assessments are presented in Supplementary Material II with Supplementary Tables 1 and 2. They show that in relation to the quality of reporting for patient characteristics and task parameters, 0 % of studies were in the first quartile, 21 % in the second quartile, 37 % in the third quartile and 42 % in the fourth quartile. In terms of the quality of neuroimaging reporting, 0 % were in the first quartile, 5 % in the second quartile, 5 % in the third quartile and 90 % in the fourth quartile. Thus the majority were scoring moderate to high levels of quality. The highest points of variability of study quality were concerning the sample sizes and characterising of participants' AVHs. Only four studies examined an AVH group with over 20 participants, with the overwhelming majority investigating AVH groups of 15 participants or less. In terms of characterising participants' AVHs, 26 of the 38 studies administered standardised clinical assessments of AVH symptomatology and of these, 16 presented the results. Of the remaining 12 studies, seven administered a standardised assessment of overall positive symptomatology, two used unstandardised, study-specific AVH assessments, two relied on participant self-report and one provided no information concerning how AVHs were assessed.

3.4. Study synthesis

The synthesis of the studies was according to the linguistic complexity of the external speech task implemented, beginning with basic speech sounds (Table 1), followed by single words (Table 2), and finally, strings of words or sentences (Table 3). Within each of these categories, studies were first organised as either clinical or non-clinical voice-hearing studies and following this, ordered, where possible, by (i) passive listening, (ii) active listening, (iii) passive listening with affectivity, (iv) AVH-mimicking affectivity and finally, (v) active listening with affectivity. Active listening was defined as any listening task which may engage attention or working memory processes, such as source identification or discrimination, and affectivity was defined as any language with either a positive or negative emotional valence.

3.4.1. Speech sounds

Overall, six studies investigated cortical function when listening to the fundamental units of spoken language and basic speech sounds (see Table 1). Included here are three investigations of clinical voice-hearers

Table 1
Studies of processing speech sounds in individuals with AVH.

Study	Sample	AVH classification	Research design	Main AVH behavioural findings[1]	Main AVH neuroimaging findings[1]
Clinical voice-hearers					
Dichotic listening (n = 2)					
Korsnes et al. (2010)	6 E-AVH 6 HC	<u>*Type:*</u> unstandardised, study-specific measure <u>*Detail:*</u> 11 questions on the phenomenology of AVH (*e.g.* personification, negative content) <u>*Scores:*</u> n.a., frequencies ranged from daily to biannually	<u>*Modality:*</u> 3 T fMRI <u>*Task:*</u> consonant-vowel dichotic listening. Interaural intensity differences & attention task <u>*Analysis:*</u> ROI (small volume corrections)	↓ R ear advantage	<u>*E-AVH v. HC:*</u> ↓ 12% L lateralisation of STG & IFG <u>*Correlations:*</u> ↑ AVH severity = ↓ accuracy, ↓ STG activation
Steinmann et al. (2017)	13 Sz-AVH 13 Sz-SN 26 HC	<u>*Type:*</u> standardised measure <u>*Detail:*</u> PANSS item P3 <u>*Scores:*</u> Sz-AVH: 4.1 ± 1.2; Sz-SN: 1.0 ± 0.0	<u>*Modality:*</u> EEG <u>*Task:*</u> consonant-vowel dichotic listening. Interaural attention task[3] <u>*Analysis:*</u> gamma band (30–100 Hz) synchrony between L – R A1. Created a laterality index based on L & R ear accuracy. Lagged phase synchronisation analysis	↓ laterality index ↑ errors	<u>*Sz-AVH v. HC; Sz-AVH v. Sz-SN:*</u> ↑ synchrony differences between L & R ear at 600 ms post-stimulus ↑ synchrony during conscious perception of L ear <u>*Correlations:*</u> *↑ PANSS item P3 = ↑ synchrony*
Actively listening to speech sounds (n = 1)					
Heinks-Maldonado et al. (2007)	10 Sz-AVH[2] 10 Sz-TN 17 HC[2]	<u>*Type:*</u> self-report & standardised measure <u>*Detail:*</u> current AVH & ≥ 2 on SAPS AVH (‘auditory hallucination’, ‘voices commenting’ and ‘voices conversing’ items) <u>*Scores:*</u> SAPS AVH: Sz-AVH: 2.1 ± 0.6; Sz-TN: 0.9 ± 0.4	<u>*Modality:*</u> EEG <u>*Task:*</u> the syllable “ah”. Self (distorted, undistorted) & alien (distorted, undistorted) conditions. Source discrimination task[3] <u>*Analysis:*</u> *N100 (80–120 ms)* amplitude & suppression measured across 20 electrode sites	↓ accuracy for self-distorted & alien-undistorted speech	<u>*Sz-AVH v. HC:*</u> *Self-undistorted voice:* ↑ L N100 <u>*Sz-AVH v. Sz-TN:*</u> No group differences <u>*Correlations:*</u> ↑ SAPS AVH = ↑ N100 suppression
Non-clinical voice-hearers					
Dichotic listening (n = 2)					
Kompus et al. (2013)	8 H-AVH 8 HC[2]	<u>*Type:*</u> self-report & standardised measure <u>*Detail:*</u> 6 childhood onset; 2 recent adult onset in the last 4 years; confirmed with LSHS & PSYRATS <u>*Scores:*</u> n.s.	<u>*Modality:*</u> 3 T fMRI <u>*Task:*</u> consonant-vowel dichotic listening. Interaural intensity differences & attention task[3] <u>*Analysis:*</u> ROI	*Auditory acuity:*↓ at 2000 Hz & 3000 Hz (both ears); ↑ at 500 Hz & 2000 Hz (L ear only)	<u>*H-AVH v. HC:*</u> ↓ R A1 activity ↑ lateral spread of A1 activation in H-AVH
Thiebes et al. (2018)	25 HC[2] (10 H-AVH) (15 H-N)	<u>*Type:*</u> standardised measure <u>*Detail:*</u> induced with intravenous ketamine.	<u>*Modality:*</u> EEG <u>*Task:*</u> consonant-vowel dichotic	None	<u>*H-AVH v. H-N:*</u> ↑ L ear gamma band connectivity

(*continued on next page*)

Table 1 (*continued*)

Study	Sample	AVH classification	Research design	Main AVH behavioural findings[1]	Main AVH neuroimaging findings[1]
		Assessed with AVH-specific version of 5D-ASC AUA & PANSS item P3 *Scores:* n.s.	listening. Interaural attention task[3] *Analysis:* N100 (Cz), baseline gamma power & auditory-evoked gamma power responses measured across 27 electrodes. Gamma connectivity (30 – 100 Hz) measured by lagged phase synchronization across L & R A1 & A2		
	Actively listening to speech sounds (n = 1)				
Pinheiro et al. (2018)	32 H-AVH[2]	*Type:* clinical interview & standardised measure *Detail:* monthly AVH, LSHS AVH *Scores:* 3.2 ± 2.9	*Modality:* EEG *Task:* auditory (passive listening), motor (button press) & auditory-motor conditions. Tones & the speech syllable "ah". Self & alien conditions[3] *Analysis:* *N100 (70–110 ms), P2 (170–210 ms)* amplitudes & pre-stimulus alpha power (8–12 Hz; −250 to 0 ms) across L & R, anterior & posterior electrode clusters	n.a.	*Correlations:* *N100 negativity:* positive correlation with LSHS total & LSHS AVH *Pre-stimulus alpha power:* ↑ LSHS total = ↓ power for tones compared to voices.

Note: Scores are presented as mean ± SD. Hz: hertz; L: left; R: right; n.a.: not applicable; n.s.: not stated.
1: all behavioural and correlational results are specific to AVH group unless otherwise specified; 2: 100 % right-handed; 3: button-press used.
Sample names: AVH: auditory verbal hallucination; E-AVH: epilepsy AVH; H-AVH: non-clinical AVH; HC: healthy controls; H-N: non-AVH group after ketamine administration; Sz: schizophrenia; Sz-AVH: schizophrenia AVH; Sz-SN: state negative: Sz group with some/minimal clinical history of AVH; Sz-TN: trait negative - Sz group with no clinical history of AVH.
Psychopathological scales: 5D-ASC: Altered States of Consciousness questionnaire; LSHS: Launay-Slade Hallucination Scale (maximum scores: total: 64; AVH: 16); PANSS: Positive and Negative Syndrome Scale; PANSS item P3: PANSS hallucinatory behaviour (maximum score: 7); PSYRATS: Psychotic Symptom Rating Scale; SAPS: Scale for the Assessment of Positive Symptoms (SAPS AVH: combined scores of SAPS items 'auditory hallucinations', 'voices commenting' and 'voices conversing'; maximum score: 15).
Cortical & neuroimaging abbreviations: A1: primary auditory cortex; aeGBR: auditory-evoked gamma band response; IFG: inferior frontal gyrus; ROI: region of interest; STG: superior temporal gyrus.

Table 2
Studies of processing single words in individuals with AVH.

Study	Sample	AVH classification	Research design	Main AVH group behavioural findings[1]	Main AVH group neuroimaging findings[1]
Clinical voice-hearers **Passively listening to single words (n = 5)**					
Gavrilescu et al. (2010)	14 Sz-AVH 13 Sz-TN 14 HC	*Type:* standardised measure *Detail:* AVH-specific modification of PANSS item P3. Current PANSS P3-AVH score ≥3 & a history of ≥3 for one month or longer *Scores:* Sz-AVH: 3.8 ± 0.7; Sz-TN: 1.4 ± 0.5	*Modality:* 3 T fMRI *Task:* monoaural & binaural presentation of neutral words, silence[3] *Analysis:* seed-based analysis - bilateral A1 & A2	n.a.	*Sz-AVH v. HC; Sz-AVH v. Sz-TN:* ↓ functional connectivity of L – R A1 and L – R A2

(*continued on next page*)

Table 2 (*continued*)

Study	Sample	AVH classification	Research design	Main AVH group behavioural findings[1]	Main AVH group neuroimaging findings[1]
Innes-Brown et al. (2006)	22 Sz-AVH 26 Sz-SN 22 HC	*Type:* self-report *Detail:* current AVH *Scores:* n.a.	*Modality:* EEG *Task:* monoaural presentation of tones or neutral words *Analysis:* temporal interhemispheric N100 amplitude (ms not defined) from 14 electrode sites	n.a.	*Sz-AVH v. HC:* R words = ↓ R N100 in Sz-AVH
Henshall et al. (2012)[a]	19 Sz-AVH 17 Sz-N 17 HC	*Type:* self-report & standardised measure *Detail:* AVH experienced in the past week, PANSS item P3 *Scores:* Sz-AVH: 3.9 ± 1.3; Sz-N: 1.9 ± 0.8	*Modality:* EEG *Task:* monoaural presentation of tones or neutral words *Analysis:* amplitude, latency & interhemispheric transfer time (IHTT) of N100 at C3, C4 (90–150 ms) & C5, C6 (90–190 ms) electrode pairs. All results here are after gender factored into analyses	n.a.	*Sz-AVH v. HC:* No differences for words *Sz-AVH v. Sz-N:* ↑ IHTT at C3-C4
Henshall et al. (2013)[a]	19 Sz-AVH 17 Sz-N 17 HC	*Type:* self-report & standardised measure *Detail:* AVH experienced in the past week, PANSS item P3 scores *Scores:* Sz-AVH: 3.9 ± 1.3; Sz-N: 1.9 ± 0.8	*Modality:* EEG *Task:* monoaural presentation of tones or neutral words *Analysis:* high alpha (10 – 12 Hz) & high beta (22 – 30 Hz) band coherence at C1-C2, C3-C4, C5-C6, T7-T8, Ft7-Ft8, Cp5-Cp6 electrode pairs. All results here are after gender factored into analyses	n.a.	*Sz-AVH v. HC:* *Alpha coherence:* ↓ C3-C4, C5-C6 *Sz-AVH v. Sz-N: Alpha coherence:* ↓ C3-C4, C5-C6 *Beta coherence:* ↓ C3-C4 *Correlations:* No effect of ear presentation or gender across all groups
Bühler et al. (2016)	14 Sz-AVH 14 Sz-N 28 HC	*Type:* medical history & standardised measure *Detail:* frequent & current AVH as per medical history, AHRS, PANSS, PSYRATS *Scores:* AHRS total: 21.7 ± 9.6;m Sz-N: 0.0; PANSS P: Sz-AVH: 16.3 ± 4.7; Sz-N: 15.3 ± 3.6'; PSYRATS total: Sz-AVH: 31.5 ± 13.7; Sz-N: 10.0 ± 7.0	*Modality:* EEG *Task:* constant pink noise. Neutral words in self or alien voice, with or without auditory delay, or passive visual fixation. Six conditions assessing agency & ownership of speech *Analysis: N100 (116*–170 ms) & late component (172–356 ms) from Fz electrode. Template-based analysis (for ERP activity), TANOVA (for spatial distribution/ topography of ERP) & global field power	No group differences in voice recognition response time	*Sz-AVH v. HC:* *Late component spatial distribution:* ↑ effect of agency *Sz-AVH v. Sz-N:* No group differences *Correlations:* No effect of AVH severity
Actively listening to single words (n = 2)					
Ganguli et al. (1997)	8 Sz-TP[2] 8 HC	*Type:* clinical interview & standardised measure *Detail:* no current AVH reported by clinicians, assessed with SADS	*Modality:* $[^{15}O]$-H_2O PET *Task:* emotionally neutral words & silence. Auditory supraspan memory task & passive visual fixation	↓ primary memory accuracy	*Sz-TP v. HC:* *Supraspan memory:* ↓ overall bilateral temporal & frontal regions. ↓ L STG, R putamen, L THAL, ↑ L OFC, L temporal

(*continued on next page*)

Table 2 (*continued*)

Study	Sample	AVH classification	Research design	Main AVH group behavioural findings[1]	Main AVH group neuroimaging findings[1]
		Scores: n.s.	*Analysis:* rCBF		occipital cortex, L PCC, L IPC *Passive visual fixation:* ↓ bilateral dlPFC, R STG *Supraspan memory v passive visual fixation:* ↓ bilateral ACC, dlPFC & STG
Ikuta et al. (2015)	8 Sz-AVH[2] 8 Sz-N[2]	*Type:* standardised measure *Detail: :* AVH-specific modification of BPRS hallucination score, ≥ 2 in the past week *Scores:* Sz-AVH: 2.3 ± 1.5; Sz-N: 1.0 ± 0.0	*Modality:* 3 T fMRI *Task:* single words, reversed words, sine wave sounds & silence. One-back task *Analysis:* partial brain	No AVH-specific differences in accuracy	*Correlations:* *Reversed words > sine wave sounds:* ↑ BPRS score = ↑ bilateral posterior basal ganglia, mostly within globus pallidus
Passively listening to affective single words (n = 1)					
Kang et al. (2009)	14 Sz-AVH[2] 14 Sz-TN[2] 28 HC[2]	*Type:* self-report & standardised measure *Detail:* daily AVH for ≥ 2 years, PANSS item P3 ≥ 3; AHS *Scores:* AHS total: Sz-AVH: 23.3 ± 4.7; Sz-TN: 0.0; PANSS item P3: Sz-AVH: 3.7 ± 0.5; Sz-TN: 1.0 ± 0.0	*Modality:* 1.5 T fMRI *Task:* laughing, crying, the sound “ah”, silence. Gender discrimination task[3] *Analysis:* whole brain (small volume corrections to AMYG)	No AVH-specific differences in accuracy	*Sz-AVH v. HC:* All emotion: ↑ bilateral precuneus, ↓ L cingulate *Laugh:* ↑L MFG & FG, bilateral MTG, R SPG & uncus, ↓ bilateral SFG, R caudate *Cry:* ↑ bilateral SFG. ↓ bilateral AMYG, R putamen, L MFG *Sz-AVH v. Sz-TN:* *All emotion:* ↑ R IPC, ↓ R SFC *Laugh:* ↑ R IFG, bilateral caudate, ↓ PCC, R putamen. *Cry:* ↓ bilateral HIP, L AMYG, IFG & insula
Passively listening to affective single words with AVH mimicking content (n = 4)					
De la Iglesia-Vaya et al. (2014)[b]	27 Sz-AVH[2] 14 Sz-TN[2] 31 HC[2]	*Type:* clinical interview & standardised measure *Detail:* daily AVH for ≥ 1 year, confirmed by psychiatrist. BPRS, PANSS & PSYRATS measured from last 24 hours *Scores:* BPRS total: Sz-AVH: 51.1 ± 10.9; Sz-TN: 38.9 ± 8.0; PANSS total: Sz-AVH: 65.3 ± 17.8; Sz-TN: 53.7 ± 11.7; PSYRATS total: Sz-AVH: 30.3 ± 4.96; Sz-TN: 0	*Modality:* 1.5 T fMRI *Task:* neutral & AVH-mimicking emotional words. Passive listening *Analysis:* independent component analysis of network synchrony, Granger causal analysis of effective connectivity	n.a.	*Sz-AVH v. HC; Sz-AVH v. Sz-TN:* *Synchrony:* ↓ occipito-cerebellar & limbic networks; ↓ fronto-temporal, temporal & temporo-parietal networks. *Causal source of effective connectivity:* Sz-AVH: occipito-cerebellar network; HC & Sz-TN: temporal network.
Escartí et al. (2010)[b]	27 Sz-AVH[2] 14 Sz-TN[2]	*Type:* clinical interview & standardised measure *Detail:* daily AVH for ≥ 1 year, confirmed by psychiatrist.	*Modality:* 1.5 T fMRI *Task:* neutral & AVH-mimicking emotional words. Passive listening	n.a.	*Sz-AVH v. HC; Sz-AVH v. Sz-TN:* ↑ AMYG, paHIP. ↓ STG, IFG, L insula

(continued on next page)

Table 2 (*continued*)

Study	Sample	AVH classification	Research design	Main AVH group behavioural findings[1]	Main AVH group neuroimaging findings[1]
	31 HC[2]	BPRS, PANSS & PSYRATS measured from last 24 hours *Scores:* BPRS total: Sz-AVH: 51.1 ± 10.9; Sz-TN: 38.9 ± 8.0; PANSS total: Sz-AVH: 65.3 ± 17.8; Sz-TN: 53.7 ± 11.7; PSYRATS total: Sz-AVH: 30.3 ± 4.96; Sz-TN: 0	*Analysis:* independent component analysis		
Sz-AVH v. HC:	Martí-Bonmatí et al. (2007)[c]				n.a.
		21 Sz-AVH[2]	*Type:* clinical interview & standardised measure	*Modality:* 1.5 T fMRI	*Emotional content:* ↑ activation of bilateral MTG, R STG, R A1, R superomedial frontal, R AG, R PCC, L MCC, R THAL
		↓ VBM density: bilateral insula, bilateral lingual gyrus, L postcentral gyrus, R precuneus, R superomedial frontal, L MTG *Detail:* persistent AVH confirmed by psychiatrist, PANSS & PSYRATS	*Task:* neutral & AVH-mimicking emotional words. Passive listening		*Correlations:*
	10 HC[2]	*Scores:* PANSS total: 71.0 ± 10.0; PSYRATS total: 30.0 ± 4.0	*Analysis:* T1 & T2 weighted whole brain		*(↑ activation + ↓ density) coincidence:* ↑ bilateral MTG & STG, ↓ L PCC, R ACC, L inferior opercularis, R MOG
Sanjuan et al. (2007)[c]				↑ anxiety induced by emotional speech	*Sz-AVH v. HC:*
	11 Sz-AVH[2]	*Type:* self-report & standardised measure	*Modality:* 1.5 T fMRI		*Emotional speech v. neutral speech:* ↑ L MTG, bilateral insula, R median cingulate, bilateral PCC, R AMYG, bilateral orbital MFC, R orbital IFG, R SMC
		Detail: daily AVH for ≥ 1 year, BPRS, PANSS, PSYRATS	*Task:* neutral & AVH-mimicking emotional words. Review task post-scanning of voice resemblance & anxiety		*Neutral speech v. silence:* ↑ L MTG, L STG, middle cingulate, L orbital IFG
	10 HC[2]	*Scores:* BPRS total: 55.2 ± 7.4; PANSS total: 70.8 ± 9.9; PSYRATS total: 28.6 ± 4.4	*Analysis:* whole brain		
Actively listening to affective single words (n = 3)					
Allen et al. (2007)[d]	10 Sz-AVH[2]	*Type:* self-report & standardised measure	*Modality:* 1.5 T fMRI	↓ accuracy for self-speech	*Sz-AVH v. HC; Sz-AVH v. Sz-TN:*
	10 Sz-TN[2]	*Detail:* prominent & current AVH, >3on SAPS ‘auditory hallucination’	*Task:* self (distorted, undistorted) & alien (distorted, undistorted) conditions. Emotional (positive & negative valence) or neutral words. Source discrimination task[3]		↓ L STG for alien > self speech
	11 HC[2]	*Scores:* Sz-AVH: 4.5 ± 0.7; Sz-TN: 0	*Analysis:* whole brain		↓ cingulate gyrus for distorted > undistorted speech

(*continued on next page*)

Table 2 (*continued*)

Study	Sample	AVH classification	Research design	Main AVH group behavioural findings[1]	Main AVH group neuroimaging findings[1]
					↑ R STG for distorted self > undistorted self speech ↓ R ACG for distorted alien > undistorted alien speech. A loss of alien distortion effect was seen in Sz-AVH *Correlations:* ↑ L MTG = ↑ accuracy in HC & Sz-TN. Not seen in Sz-AVH
Mechelli et al. (2007)[d]	11 Sz-AVH[2] 10 Sz-TN[2] 10 HC[2]	*Type:* self-report & standardised measure *Detail:* prominent & current AVH, >3 on SAPS 'auditory hallucination' *Scores:* Sz-AVH: 4.4 ± 0.7; Sz-TN: 0	*Modality:* 1.5 T fMRI *Task:* self (distorted, undistorted) & alien (distorted, undistorted) conditions. Emotional (positive & negative valence) & neutral adjectives. Source discrimination task[3] *Analysis:* partial brain, dynamic causal modelling	↓ accuracy for self-speech	*Sz-AVH v. HC:* *Intrinsic connections:*↓ bilateral STG – ACC *Distorted > undistorted speech:* ↓ bilateral STG – ACC *Sz-AVH v. Sz-TN:* No group differences *Effect of speech source:* Self-speech = ↑ intrinsic connections inSz-AVH Alien speech =↑ intrinsic connections in HC & Sz-TN
Pinheiro et al. (2017)	15 SSD-AVH[2] 16 HC[2]	*Type:* standardised measure *Detail:* PANSS & SAPS, confirmed with personal communication *Scores:* PANSS P: 23.0 ± 9.2 SAPS global: 11.0 ± 3.4	*Modality:* EEG *Task:* self & alien speech. Emotional (positive & negative valence) & neutral words. Source discrimination task[3] *Analysis: N100 (130–210 ms), P2 (HC: 215–380 ms, Sz-AVH: 250*–415 ms), late positive potential (LPP; 500–700 ms) amplitudes across frontocentral, central & centroparietal electrodes	↓ recognition accuracy of negative speech source	*SSD-AVH v. HC:* *N100*: no group differences *P2:* no group differences *LPP:* SSD-AVH differences during *negative* speech (self > alien), whereas in HC, all self > alien, regardless of emotional valence *Correlations:* ↑ PANSS hallucination severity = ↑ LPP amplitude for negative alien speech ↑ SAPS 'voices conversing' scores = ↑ difference between LPP responses to negative v. positive emotional valence
Non-clinical voice-hearers **Actively listening to single words (n = 1)**					
Lewis-Hanna et al. (2011)	12 H-AVH[2] 12 HC[2]	*Type:* unstandardised, study-specific measure *Detail:* 2 questions on the phenomenology of hypnogogic/ hypnopompic AVH *Scores:* n.a., frequencies ranged from weekly – only heard once	*Modality:* 1.5 T fMRI *Task:* two tasks: (i) passive listening to neutral words; (ii) mismatched presentation of visual & aural numbers during a forced auditory or visual attention task *Analysis:* VOI	↑ auditory acuity No AVH-specific differences in accuracy or reaction time	*H-AVH v. HC:* *Passive listening > silence:* ↑ L posterior TPC, L SMG *Auditory > visual attention:* ↑ R ACC *Correlations:* ↑ auditory acuity = ↑ L SMG across both groups

Note: Scores are presented as mean ± SD. [a, b, c, d] denote overlapping samples. L: left; R: right. n.a.: not applicable; n.s.: not stated.
1: all behavioural and correlational results are specific to AVH group unless otherwise specified; 2: 100 % right-handed; 3: button-press used; ‘: differences in AVH or positive symptomology severity between Sz groups was insignificant.
Sample names: AVH: auditory verbal hallucination; H-AVH: non-clinical AVH; HC: healthy controls; IHTT: interhemispheric transfer time; Navh: non-AVH group; rCBF: regional cerebral blood flow; SSD: schizophrenia spectrum disorder; Sz: schizophrenia; Sz-AVH: schizophrenia AVH; Sz-N: Sz non-AVH group with either mixed or unspecified history of AVH; Sz-SN: state negative: Sz group with some/minimal clinical history of AVH; Sz-TN: trait negative - Sz group with no clinical history of AVH; Sz-TP: trait positive – history of SzAVH & not actively hallucinating/in symptomatic remission.
Psychopathological scales: BPRS: Brief Psychiatric Rating Scale (maximum auditory hallucination score: 7); DSM-III-R: 1987 revision of the Diagnostic and Statistical Manual of Mental Disorders; MMSE: Mini Mental State Examination (maximum score: 30); PANSS: Positive and Negative Syndrome Scale PANSS P: PANSS positive (maximum score: 49); PANSS item P3: PANSS hallucinatory behaviour (maximum score: 7); PSYRATS: Psychotic Symptom Rating Scale (maximum score: 44); SADS: Schedule for Affective Disorders & Schizophrenia Interview; SAPS: Scale for the Assessment of Positive Symptoms (maximum scores: global: 20; AVH: 15); SSPI: Signs and Symptoms of Psychotic Illness.
Cortical & neuroimaging abbreviations: [^{15}O]-H_2O: 15-Oxygen labelled water; A1: primary auditory cortex; A2: secondary auditory cortex; ACC: anterior cingulate cortex; ACG: anterior cingulate gyrus; AG: angular gyrus; AMYG: amygdala; dlPFC: dorsolateral prefrontal cortex; FG: fusiform gyrus; HIP: hippocampus; IFG: inferior frontal gyrus; IPC: inferior parietal cortex; MFC: middle frontal cortex; MFG: middle frontal gyrus; MPC: middle parietal cortex; MOG: middle occipital gyrus; MTG: middle temporal gyrus; OFC: orbitofrontal cortex; paHIP: parahippocampal gyrus; PCC: posterior cingulate cortex; PFC: prefrontal cortex; rCBF: regional cerebral blood flow; SFC: superior frontal cortex; SMC: superior medial cortex; SMG: supramarginal gyrus; STG: superior temporal gyrus; THAL: thalamus; TPC; temporoparietal cortex; VBM: voxel based morphometry; VOI: voxel of interest.

Table 3
Studies of processing strings of words or sentences in individuals with AVH.

Study	Sample	AVH classification	Research design	Main AVH behavioural findings[1]	Main AVH neuroimaging findings[1]
Clinical voice-hearers					
Passively listening to sentences (n = 7)					
Briend et al. (2017)	11 Sz-AVH 10 HC	*Type:* standardised measure *Detail:* AHRS evaluation *Scores:* n.s.	*Modality:* 3 T fMRI *Task:* 20 Hz rTMS over L STS, only on Sz-AVH cohort. Neutral story *versus* silence *Analysis:* seed-based functional connectivity of bilateral A1 & TG	n.a.	*Sz-AVH v. HC:* Baseline: ↓ functional connectivity between L – R A1, L – R TG, and R A1 – R TG *Post-rTMS:* no functional connectivity change, ↓ AHRS score
Jardri et al., (2011a)	15 Sz-AVH[2] 15 HC[2]	*Type:* standardised measure *Detail:* PANSS, confirmed by personal communication with authors *Scores:* PANSS P: 22.5 ± 5.5	*Modality:* 1.5 T fMRI *Task:* self-speech listening paired with inner repetition, alien & reversed alien speech both passive listening. Used a poem (emotional valence unknown). Discrimination task *Analysis:* partial brain	No group differences in accuracy	*Sz-AVH v. HC:* *Intelligible language effects (alien speech - reversed alien speech):* ↓ R MNS, R MTG after controlling for education *Self-agency effects (alien speech - self speech):* ↓ R IPL, R MTG, MeFG, MePC, ACC, PCC after controlling for education ↓ voxel cluster sizes of R IPL, R MTG, MeFG & MePC *Correlations:* ↑ PANSS P = ↑ R IPL
Lavigne et al. (2015)[a]	10 Sz-AVH[2] 13 Sz-SN	*Type:* clinical interview & standardised measure *Detail:* clinical interview to confirm AVH presence,	*Modality:* 3 T fMRI *Task:* passive listening to definitions of	n.a.	*All groups*: Isolated hemodynamic response (HDR)

(*continued on next page*)

Table 3 (*continued*)

Study	Sample	AVH classification	Research design	Main AVH behavioural findings[1]	Main AVH neuroimaging findings[1]
	22 BD	SSPI hallucination score ≥ 3	common words paired with corresponding images		shape of two networks during listening (results are of these combined): (i) language processing: ↑ L p. opercularis (L IFG), bilateral STG, FG, visual & SMA; (ii) default mode: ↑ bilateral visual & FG, ↓ PCC, precuneus, mPFC, SFC, IPC, LOC, PGC, SPC
Sz-AVH v. all groups:					
		Scores: Sz-AVH: 3.7 ± 0.5; Sz-SN: 0.5 ± 0.8; BD: 0.1 ± 0.4	*Analysis:* constrained principal component analysis	27 HC	↑ HDR of both networks between 0–7.5 & 20–22.5 s post-stimulus *Sz-AVH v. Sz-SN:* ↑ HDR of both networks between 0–5 & 15–22.5 s post-stimulus *Sz-AVH v. BD:* ↑ HDR of both networks between 0–7.5 & 15–20 s post-stimulus *Correlations:*
		↑ SSPI hallucination score = ↑ HDR of all networks			
Lavigne and Woodward (2018)[a]	12 Sz-AVH	*Type:* clinical interview & standardised measure	*Modality:* 3 T fMRI	n.a.	*All groups*:
	11 Sz-SN	*Detail:* clinical interview to confirm AVH presence, SSPI hallucination score ≥ 2	*Task:* passive listening to definitions of common words paired with corresponding images		Isolated HDR shape of three networks during listening (results are of these combined): (i) auditory-motor network: ↑ bilateral TP, STG, SMA, dACC, VC, insula, THAL, cerebellum, L PCG; (ii) language processing network: ↑ L pMTG, IFG, OFC, dlPFC, VC; (iii) default mode:↑ bilateral STG, VC, L PCG, SMA, A1, ↓ bilateral vmPFC, precuneus, PCC, lateral OCC *Sz-AVH v. HC:* ↑ HDR of all networks 3.75, 6.25 & 8.75 s post-stimulus
	27 HC	*Scores:* n.s.	*Analysis:* constrained principal component analysis		*Sz-AVH v. Sz-SN:* ↑ HDR of all networks 6.25 & 8.75 s post-stimulus
Rapin et al (2012)	5 SSD-AVH[2]	*Type:* self-report & standardised measure	*Modality:* 3 T fMRI	n.a.	*SSD-AVH & HC:*
	10 HC[2]	*Detail:* one participant reported AVH in the past week. Confirmed with SSPI & personal communication with authors	*Task:* passive listening to definitions of common words paired with corresponding images		Isolated HDR shape of one network during listening: ↑ bilateral STG, lingual/FG/occipital/cerebellar regions, ↓ bilateral superior frontal regions, angular gyrus, SMG

(*continued on next page*)

Table 3 (*continued*)

Study	Sample	AVH classification	Research design	Main AVH behavioural findings[1]	Main AVH neuroimaging findings[1]
		Scores: SSPI total: 11.0 ± 6.0	*Analysis:* constrained principal component analysis		*SSD-AVH v. HC:* ↑ HDR of this network between 9 – 18 s post-stimulus
Woodruff et al. (1997)	7 state positive (SP-) & state negative (SN-) AVH[2]	*Type:* standardised measure	*Modality:* 1.5 T fMRI	n.a.	*SP-AVH v. SN-AVH:* ↓ R MTG, L STG
	8 Sz-TP[2]	*Detail:* as per SAPS 'auditory hallucination' scores: SP-AVH: severe, ongoing, current Sz-AVH; SN-AVH: diminished Sz-AVH symptomology; Sz-TP: self-reported history of AVH, not actively hallucinating	*Task:* neutral story *versus* silence		*Sz-TP v. Sz-TN:*
	7 Sz-TN[2] 8 HC[2]	*Scores:* SP-AVH: 4.7 ± 0.5; SN-AVH: 2.7 ± 2.1; r-Sz-AVH: 3.0 ± 2.5; Sz-TN: 0	*Analysis:* fronto-temporo-occipital		No group differences
Zhang et al. (2008b)[b]	13 FE/SE-AVH[2]	*Type:* standardised measure	*Modality:* 1.5 T fMRI	n.a.	*FE/SE-AVH v. HC:* R voice = ↓ R MFG
	13 FE/SE-TN[2]	*Detail:* SAPS 'auditory hallucination' score ≥ 4	*Task:* monoaural presentation of familiar (*e.g.* participants' family members) & unfamiliar voices. Commanding sentences		*FE/SE-AVH v. FE/SE-TN:*
	13 HC[2]	*Scores:* n.s.; SAPS total: FE/SE-AVH: 45.5 ± 13.2; FE/SE-TN: 22.2 ± 6.7	*Analysis:* whole brain		L & R voice = ↑ L Wernicke's, SMG, AG & STG
Actively listening to sentences (n = 4)					
Mou et al. (2013)[b]				↓ voice recognition accuracy	*FE/SE-AVH v. HC:*
	13 FE/SE-AVH[2]	*Type:* standardised measure	*Modality:* 1.5 T fMRI		↓ functional connectivity of R STG – R SFG & R STG – L MFG
	13 FE/SE-TN[2]	*Detail:* SAPS 'auditory hallucination' score ≥ 4	*Task:* familiar (*e.g.* participants' family members) & unfamiliar voices. Commanding sentences. Source discrimination task[3]		*FE/SE-AVH v. FE/SE-TN:*
	13 HC[2]	*Scores:* n.s.; SAPS total: FE/SE-AVH: 45.5 ± 13.2; FE/SE-TN: 22.2 ± 6.7	*Analysis:* seed-based functional connectivity of R STG		↓ functional connectivity of R STG – R SFG *Correlations:* *FE/SE-AVH functional connectivity:* ↑ R STG – R SFG = ↑ voice recognition accuracy
Plaze et al. (2006)		*Type:* self-report & standardised measure	*Modality:* 1.5 T fMRI	85% accuracy in sentence detection	*Speech > silence:*
	15 Sz-AVH[2]	*Detail:* daily AVH for minimum 3 months, AVH subscales of PSYRATS (11 items) & SAPS ('auditory hallucination', 'voices commenting' and 'voices conversing' items)	*Task:* neutral sentences *versus* silence. Sentence-fragment confirmation task[3]		↑ bilateral STS, IFG
		Scores: PSYRATS AVH: 30.0 ± 6.0 SAPS AVH: 10.0 ± 3.0	*Analysis:* whole brain. Small volume corrections for SAPS AVH regression		*Correlations:* ↑ PSYRATS = ↓ posterior L STG ↑ SAPS AVH = ↓ posterior L STG
Stephane et al. (2018)	7 Sz-AVH 8 HC	*Type:* n.s.	*Modality:* 3 T fMRI	n.a.	*Sz-AVH v. HC:*

(continued on next page)

Table 3 (*continued*)

Study	Sample	AVH classification	Research design	Main AVH behavioural findings[1]	Main AVH neuroimaging findings[1]
		<u>*Detail:*</u> confirmed by personal communication with authors <u>*Scores:*</u> n.s.	<u>*Task:*</u> self & alien sentences. Source discrimination task <u>*Analysis:*</u> subject-level, general linear modelling		↑ ACC, SMA, insula: during *self* condition in Sz-AVH, during *alien* condition in HC
Zhang et al. (2008a)[b]	13 FE/SE-AVH[2] 13 FE/SE-TN[2] 13 HC[2]	*Type:* standardised measure *Detail:* SAPS 'auditory hallucination' score ≥ 4 *Scores* n.s.; SAPS total: FE/SE-AVH: 45.5 ± 13.2; FE/SE-TN: 22.2 ± 6.7	<u>*Modality:*</u> 1.5 T fMRI <u>*Task:*</u> monoaural presentation of familiar (*e.g.* participants' family members) & unfamiliar voices. Commanding sentences. Source discrimination task[3] <u>*Analysis:*</u> whole brain	↓ source discrimination accuracy for familiar voices compared to HC	<u>*FE/SE-AVH v. HC:*</u> *Familiar > unfamiliar voice:* ↓ R STG <u>*FE/SE-AVH v. FE/SE-TN:*</u> No group differences
Passively listening to affective sentences with AVH mimicking content (n = 3)					
Ford et al. (2002)	7 Sz-AVH 5 Sz-N 10 HC[2]	<u>*Type:*</u> clinical interview & standardised measure <u>*Detail: :*</u> clinical interview to confirm AVH presence, BPRS 'hallucinatory behaviour' score ≥ 5, SAPS <u>*Scores:*</u> n.s.	<u>*Modality:*</u> EEG <u>*Task:*</u> AVH-mimicking statements, alongside a continuous probe of the syllable "ba", broadband noise & a visual checkerboard. Talking and passive listening conditions <u>*Analysis:*</u> coherence of delta, theta, alpha, beta & gamma bands for frontal & temporal regions, alongside theta power (4–7 Hz) across 35 electrodes	n.a.	<u>*Sz-AVH v. HC:*</u> ↓ theta coherence during talking <u>*Sz-AVH v. Sz-N:*</u> ↓ theta coherence during talking
Haesebaert et al. (2013)	6 Sz-AVH[2] 12 HC[2]	<u>*Type:*</u> self-report <u>*Detail:*</u> daily AVH for ≥ 6 weeks <u>*Scores:*</u> n.a.	<u>*Modality:*</u> MEG <u>*Task:*</u> AVH-mimicking statements, the syllable "da", white noise. Passive listening, vocal repetition, inner-speech repetition & rest conditions <u>*Analysis:*</u> M100 measured ~90–150 ms post stimulus in bilateral temporal regions across 30 sensors	n.a.	<u>*Sz-AVH v. HC:*</u> ↓ L M100 in temporal regions for passive listening and inner speech of AVH mimics, no group differences for white noise or the syllable "da"

(*continued on next page*)

Table 3 (*continued*)

Study	Sample	AVH classification	Research design	Main AVH behavioural findings[1]	Main AVH neuroimaging findings[1]
Horga et al. (2014a)	9 Sz-TP[2] 8 HC[2]	*Type:* self-report & standardised measure *Detail:* AVH during first Sz episode. As per PANSS item P3, no AVH after 4 weeks of treatment with risperidone; PSYRATS *Scores:* PANSS P3: 4.9 ± 0.9 PSYRATS Frequency: 2.9 ± 0.8	*Modality:* ^{18}F-FDG PET *Task:* verbal stimuli mimicked each participants' self-reported AVH content. Predominantly derogatory, aversive language. Passive listening & rest conditions *Analysis:* rGMR	n.a.	*r-Sz-AVH v. HC:* ↑ bilateral HIP, THAL, AMYG, L OFC, R STG, brainstem, cerebellar vermis, ↓ FG *Correlations:* bilateral AMYG = ↑ bilateral A1 ↑ L AMYG = ↑ posterior THAL & medial geniculate nucleus, HIP, ↓ medial PFC, precuneus, R MTG
Actively listening to affective sentences with AVH mimicking content (n = 1)					
Horga et al. (2014b)	10 SSD-AVH 10 HC[2]	*Type:* self-report & standardised measure *Detail:* frequent daily AVH, PANSS & PSYRATS *Scores:* PANSS item P3: 5.3 ± 0.2 PSYRATS AVH 24.7 ± 1.3	*Modality:* 3 T fMRI *Task:* AVH-mimicking speech, non-speech stimuli, silence. Decision making task[3]; confirmation of speech presence per trial to establish prediction (P+) & predictive error (P-) signals *Analysis:* whole brain	No group differences in accuracy	*SSD-AVH v. HC:* ↑ R A1 activity during silence ↓ P + magnitude in R STS & MTG ↓ ACC activation differences between speech v silence ↓ P- magnitude in bilateral THAL & HIP, R paHIP & FG, L VS *Correlations:* *↑ AVH during scanning = ↑ PSYRATS Frequency, ↑ L STS, ↓ posterior THAL & VTA* ↑ PSYRATS AVH: R A1 = ↓ P-, ↑ activity during silence ↑ medication dosage = more normal P- in A1 ↓ P- magnitude of R STS & MTG = ↓ VBM density across both groups
Non-clinical voice-hearers					
Actively listening to sentences (n = 1)					
Alderson-Day et al. (2017)	12 H-AVH 17 HC	*Type:* clinical interview & standardised measure *Detail:* during interview, endorsed one of three LSHS AVH questions; PANSS, PSYRATS *Scores:* LSHS n.s.; PSYRATS AVH: 13.2 ± 4.4; PANSS P3: 4.0 ± 0.6	*Modality:* 1.5 T fMRI *Task:* intelligible & unintelligible sine-wave sentences. Discrimination task of speech amongst sound[3] *Analysis:* whole brain	↓ speech recognition response time	*H-AVH v. HC:* ↑ rostral ACC, pre-SMA, middle cingulate, L SFG *Correlations:* ↓ recognition response time = ↑ PSYRATS physical characteristics score, ↑ middle cingulate & parietal regions, ↓ mePFC

Note: Scores are presented as mean ± SD. [a, b] denote overlapping samples. L: left; R: right; n.a.: not applicable; n.s.: not stated.
1: all behavioural and correlational results are specific to AVH group unless otherwise specified; 2: 100 % right-handed; 3: button-press used.
Sample names: AVH: auditory verbal hallucination; BD: bipolar disorder; FE/SE: first or second episode psychosis; H-AVH: non-clinical AVH; HC: healthy controls; SSD: schizophrenia spectrum disorder; SP-AVH: state-positive AVH – heightened symptomology; SN-AVH: state-negative AVH – diminished AVH symptomology; Sz: schizophrenia; Sz-AVH: schizophrenia AVH; Sz-N: Sz non-AVH group with either mixed or unspecified history of AVH; Sz-SN: state negative: Sz group with some/minimal clinical history of AVH; Sz-TN: trait negative - Sz group with no clinical history of AVH; Sz-TP: trait positive – history of SzAVH & not actively hallucinating/in symptomatic remission.
Psychopathological scales: AHRS: Auditory Hallucinations Rating Scale; BPRS: Brief Psychiatric Rating Scale (maximum scores: total; 126; hallucinations: 7); LSHS: Launay-Slade Hallucination Scale; MMSE: Mini Mental State Examination; PANSS: Positive and Negative Syndrome Scale; PANSS P: PANSS positive (maximum score: 49); PANSS item P3: PANSS hallucinatory behaviour (maximum score: 7) ; PSYRATS: Psychotic Symptom Rating Scale (maximum scores: AVH: 44; Frequency: 4); SAPS: Scale for the Assessment of Positive Symptoms (maximum scores: total: 170; auditory hallucinations: 5; AVH: 15); SSPI: Signs and Symptoms of Psychotic Illness (maximum scores: total: 80; hallucination: 4).
Cortical & neuroimaging abbreviations: ^{18}F-FDG: 2-deoxy-2-[fluorine-18]fluoro-D-glucose (^{18}F-FDG); A1: primary auditory cortex; ACC: anterior cingulate cortex; AG: angular gyrus; AMYG: amygdala; dlPFC: dorsolateral prefrontal cortex; FG: fusiform gyrus; HDR: hemodynamic response; HIP: hippocampus; IFG: inferior frontal gyrus; IPC: inferior parietal cortex; IPL: inferior parietal lobe; MCC: middle cingulate cortex; MeFG: medial frontal gyrus; MePC: medial parietal cortex; MFG: middle frontal gyrus; MNS: mirror neuron system; MTG: middle temporal gyrus; OFC: orbitofrontal cortex; paHIP: parahippocampal gyrus; PCC: posterior cingulate cortex; PFC: prefrontal cortex; PCG: precentral gyrus; rGMR: relative glucose metabolism rate; rTMS: repetitive transcranial magnetic stimulation; SFC: superior frontal cortex; SMA: supplementary motor area; SMG: supramarginal gyrus; SPC: superior parietal cortex; STG: superior temporal gyrus; STS: superior temporal sulcus; THAL: thalamus; TG: temporal gyrus; VBM: voxel based morphometry; vmPFC: ventromedial prefrontal cortex; VS: ventral striatum; VTA: ventral tegmental area.

and three of non-clinical voice-hearers. Four studies, evenly divided between fMRI and EEG, administered dichotic listening tasks. With fMRI, changes to the hemispheric lateralisation of auditory regions was observed in voice-hearers (Kompus et al., 2013; Korsnes et al., 2010), while changes to interhemispheric gamma band synchrony were observed with EEG (Steinmann et al., 2017; Thiebes et al., 2018). The remaining two studies administered study-specific tasks including active listening with EEG and found positive correlations between AVH severity and N100 suppression (Heinks-Maldonado et al., 2007; Pinheiro et al., 2018).

3.4.1.1. Clinical studies of speech sound processing

3.4.1.1.1. Clinical studies of dichotic listening. Korsnes et al. (2010) examined dichotic listening with consonant-vowel syllables (*e.g.* “ba”, “da”) in six voice-hearers with epilepsy and six HC with fMRI. Epileptic seizures predominantly originated from the left temporal lobe, with four participants having previously undergone surgical resections of frontal or temporal regions. Decreased right ear advantage was observed in the voice-hearers. Using an ROI approach with small volume corrections, analyses were restricted to the ACC, inferior frontal gyrus (IFG), median and paracingulate cortices and superior temporal gyrus (STG). Voice-hearers showed a 16 % decrease in left lateralisation of STG and IFG activity during dichotic listening compared to HC. Furthermore, increasing AVH severity negatively correlated with both syllable identification accuracy and bilateral STG activation.

Examining the same dichotic listening paradigm as above with EEG, Steinmann et al. (2017) investigated gamma synchrony in 13 Sz voice-hearers, 13 Sz state-negative voice-hearers and 26 HC. Decreased right ear advantage and syllable identification accuracy were observed in the voice-hearers compared to both non-AVH groups, which did not differ. As per a lagged phase synchronization analysis and source estimation techniques, AVH-specific differences in gamma synchrony of bilateral A1 and the secondary auditory cortex (A2) were largest at 600 ms post-stimulus onset between left and right ear presentation. Gamma synchrony increased with hallucinatory behaviour scores on the Positive and Negative Symptom Scale (PANSS). Unlike both non-AVH groups, voice-hearers did not show increased interhemispheric gamma synchrony during left, compared to right, ear presentation.

3.4.1.1.2. Clinical studies of actively listening to speech sounds. The influence of source discrimination with distortion on cortical function was investigated by Heinks-Maldonado et al. (2007) in 10 voice-hearers with Sz, 10 individuals with trait-negative Sz and 17 HC with EEG. During scanning, the phoneme “ah” was presented in either their own voice (self-speech) or a voice alien to them (alien speech) with or without pitch distortion. Voice-hearers showed poorer discrimination of speech source than both non-AVH groups, misattributing their own distorted speech as being alien and undistorted alien speech as being their own. N100 amplitude and suppression of 10 electrode pairs were collapsed into overall ‘left’ and ‘right’ values. Suppression of N100 in the left hemisphere was positively correlated with both AVH scores on the Scale for the Assessment for Positive Symptoms (SAPS) and misattribution errors.

3.4.1.2. Non-clinical studies of speech sound processing

3.4.1.2.1. Non-clinical studies of dichotic listening. The two non-clinical studies of dichotic listening both implemented the same consonant-vowel paradigm described above. Kompus et al. (2013) examined the auditory acuity and dichotic listening performance of eight non-clinical voice hearers and eight HC. Decreased right ear auditory acuity at 2000 Hz and a trend towards decreased syllable identification accuracy were recorded in the voice-hearers. fMRI analysis, restricted to subdivisions of the primary auditory cortex (A1), revealed decreased activity and a more lateral spread of right A1 voxels in voice-hearers compared to healthy controls. Comparable bilateral attenuation of A1 activity was recorded between groups during forced interaural attention.

In another study using a similar task, Thiebes et al. (2018) examined 19 healthy individuals who had been intravenously administered ketamine to investigate an analogue model of AVH with EEG. With an AVH-specific subscore of the 5-Dimensional Altered States of Consciousness Rating Scale (5D-ASC), the sample was split into 10 voice-hearers and nine non-voice hearers after ketamine administration. Auditory-evoked interhemispheric gamma band activity and N100 were averaged across all electrodes and investigated with a lagged phase synchronization analysis, while baseline gamma power was measured from 27 electrode sites. AVH-specific increases in interhemispheric gamma synchrony during left ear presentation were detected.

3.4.1.2.2. Non-clinical studies of actively listening to speech sounds. The effects of self-generated phonemes in 32 non-clinical voice-hearers were investigated with EEG by Pinheiro et al. (2018). Each participant spoke the phoneme “ah”, and the vowel sound /a/ was isolated and played back during three conditions: self-generated (a button-press elicited the sound); externally-generated (the sound was presented without a button-press) and; motor (a button-press with no resulting sound; control condition). Pre-stimulus alpha power and auditory-evoked N100 and P2 responses were examined across left, right, anterior and posterior electrode clusters. Worse AVH severity, as per the Launay-Slade Hallucination Scale (LSHS) was associated with increased N100 responses for self-generated voice. Greater total scores on the LSHS were associated with decreased pre-stimulus alpha power and N100 responses for externally-generated, compared to self-generated, voice. Overall, smaller N100 and P2 amplitudes and increased pre-stimulus alpha power were seen for self-generated, compared to externally-generated, voices.

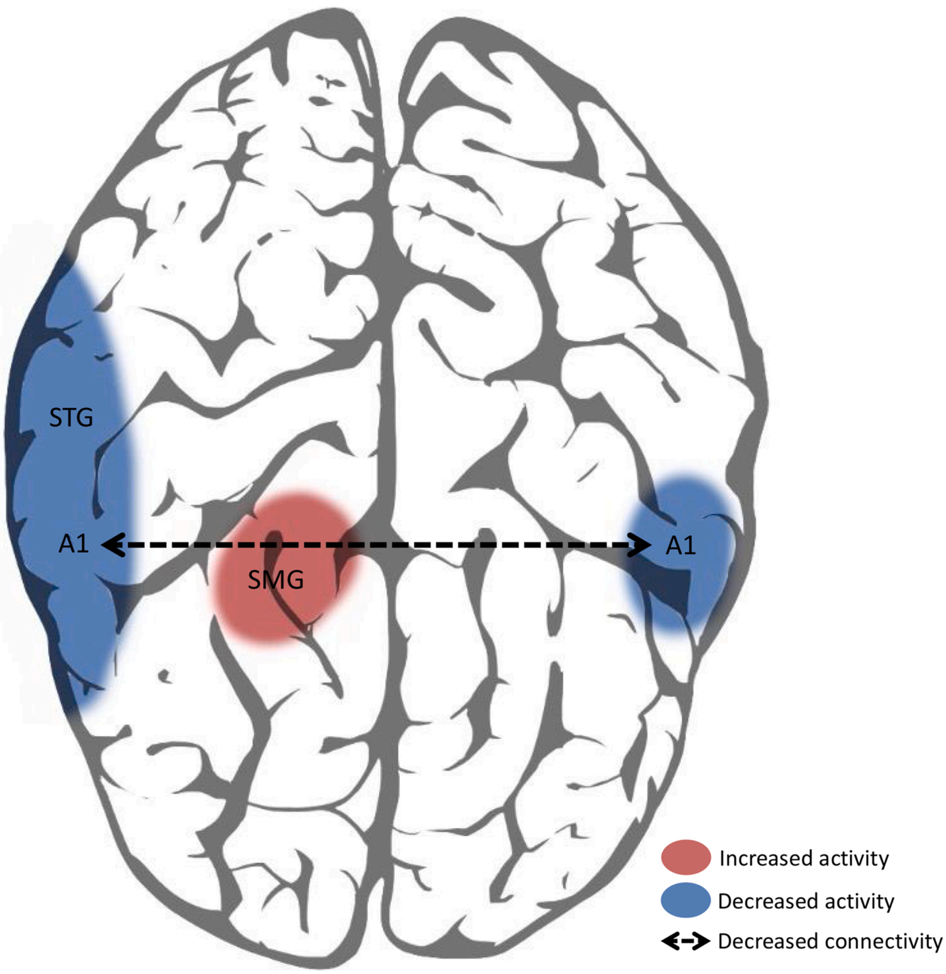

Fig. 2. Schematic representation of the key regions that have been implicated when trait voice-hearers are passively listening to external speech within functional neuroimaging investigations. Key findings during passive listening are decreased activity of the left superior temporal gyrus (STG)[1] and right primary auditory cortex (A1)[2], increased activity of the left supramarginal gyrus (SMG)[3], and decreased functional connectivity between the A1 homologues[4].
1: Allen et al. (2007), Ganguli et al. (1997), Kompus et al. (2013), Korsnes et al. (2010), Plaze et al. (2006), Woodruff et al. (1997).
2: Horga et al. (2014b), Kompus et al. (2013), Martí-Bonmatí et al. (2007), Zhang et al. (2008a).
3: Lewis-Hanna et al. (2011), Rapin et al. (2012), Zhang et al. (2008b).
4: Briend et al. (2017), Gavrilescu et al. (2010), Henshall et al. (2012), Steinmann et al. (2017).

3.4.2. Single words

Sixteen studies investigated the processing of externally presented words (see Table 2). Fifteen studies investigated Sz voice-hearers and the remaining study investigated non-clinical voice-hearers. Included in this section are 15 investigations of single words, with the final study investigating meaningful inflection of spoken sounds. Among these were five passive listening studies, three studies with active listening, one with affective content, four with affective, AVH-mimicking content and three including both affective content and active listening. Ten studies investigated cortical function with fMRI, five with EEG and one with SPECT. Common to many studies was AVH-specific differences to the activity of temporal auditory processing regions. With active listening tasks, activation differences were observed in further regions within the default mode, language and salience networks, while deviant cerebellar and limbic activity was observed when affective content was presented.

3.4.2.1. Clinical studies of single word processing

3.4.2.1.1. Clinical studies of passively listening to single words. With fMRI, Gavrilescu et al. (2010) examined the monoaural and binaural processing of neutral words and silence amongst 14 voice-hearers with Sz, 14 individuals with trait-negative Sz and 13 HC. Seed-based analysis revealed decreased connectivity between the bilateral A1 and A2 homologues in voice-hearers, both during listening and in the resting state, compared to the two non-AVH groups. No differences between monoaural and biaural presentation were reported.

The remainder of studies investigated cortical activity with EEG. Henshall et al. (2012, 2013) investigated the same sample of 19 Sz voice-hearers, 17 individuals with Sz who had not hallucinated in the past year (although it was unclear if these individuals were state- or trait-negative for AVH) and 17 HC. Words were presented monoaurally during scanning. After controlling for gender, Henshall et al. (2012) found larger N100 interhemispheric transfer time (IHTT) between C3-C4 electrodes in Sz voice-hearers compared to Sz non-voice hearers, whereas larger IHTT of C5-C6 was observed in both Sz voice-hearers and healthy controls compared to Sz non-voice-hearers. No AVH-specific differences in N100 amplitude or latency were reported. When investigating EEG coherence, Henshall et al. (2013) detected an AVH-specific loss of high alpha coherence (10–12 Hz) at electrode pairs C5-C6 (where the other groups were equal) and at C3-C4 (where values were largest in non-voice hearers with Sz, followed by HC). Further, AVH-specific losses of high beta coherence (22–30 Hz) were observed at C3-C4 compared to non-voice hearers with Sz. At the Ft7-Ft8 electrode pair, both Sz groups showed an equal loss in high alpha coherence compared to healthy controls. Neither study found any AVH-specific effects of ear of presentation.

Auditory N100 amplitude in response to monoaural and binaural presentation of words was was measured from 14 temporal electrode sites in 22 Sz voice-hearers, 26 Sz state-negative voice-hearers and 22 HC by Innes-Brown et al. (2006). During right ear presentation of words, decreased N100 amplitude in right temporal regions was specific to voice-hearers, with the two non-AVH groups showing similar responses. Comparatively, a sample of 14 Sz voice-hearers, 14 individuals with mixed state- and trait-negative Sz and 28 HC were examined by Bühler et al. (2016) when investigating the effects of speech agency and ownership. Agency conditions involved hearing either a self- or alien speech recording immediately after talking (*i.e.* being the catalyst of speech), whereas ownership was defined as hearing self-speech after talking with or without a 200 ms pause. Constant pink noise (low frequency fluctuations of sound, similar to white or brown noise) was used throughout the EEG recording. N100, measured from 116 to 170 ms post-stimulus at Cz, showed no AVH-specfic differences. A positive ERP component measured from 172 to 356 ms post-stimulus was observed as having a larger spatial distribution in voice-hearers with Sz compared to HC during conditions of agency.

3.4.2.1.2. Clinical studies of actively listening to single words. During fMRI scanning, Ikuta et al. (2015) investigated intelligible and unintelligible words in a one-back task in 16 Sz voice-hearers. The sole significant finding was in response to unintelligible words, where cortical activity of the bilateral globus pallidus within the basal ganglia positively correlated with AVH-specific scores on the Brief Psychiatric Rating Scale (BPRS).

With a $[^{15}O]$-H_2O PET scan, Ganguli et al. (1997) examined auditory supraspan memory and passive visual fixation in eight Sz individuals who previously heard voices (trait-positive, state-negative) and eight HC. Words were presented to participants either prior to or during scanning to examine working and long term memory respectively. Voice-hearers showed significantly poorer primary memory accuracy than HC, with no differences in long term memory. Compared to HC, a lack of rCBF changes in bilateral frontal and temporal areas was observed in voice-hearers during supraspan memory. Specifically, a lack of increased left STG, thalamus and right putamen rCBF, and of decreased left inferior parietal cortex, orbitofrontal cortex and temporal occipital cortex and posterior cingulate cortex (PCC) rCBF was present in voice-hearers, whereas this was observed in HC. During passive visual fixation, voice-hearers showed larger decreases in bilateral dorsolateral prefrontal cortex (dlPFC) and right STG rCBF. When comparing the two conditions, an overall decrease of bilateral ACC, dlPFC and STG rCBF was observed in the voice-hearers compared to HC.

3.4.2.1.3. Clinical studies of passively listening to affective single words. To investigate the effects of meaningful, affective vocal inflections, Kang et al. (2009) assessed cortical activity with fMRI in response to sounds of laughing and crying, compared ot the neutral sound “ah”, in 14 Sz voice-hearers, 14 individuals with trait-negative Sz and 28 HC. Numerous AVH-specific differences were observed during a whole brain analysis. Both laughing and crying elicited higher activity of right IPC and decreased activity of right superior frontal cortex (SFC) in voice-hearers compared to trait-negative Sz, alongside increased bilateral precuneus and decreased left cingulate activity, compared to HC. When specifically looking at cortical responses to laughing or crying, further differences were reported. When voice-hearers listened to laughing, increased activity of the caudate bilaterally and right IFG, and decreased activity of the PCC and right putamen was observed compared to trait-negative Sz. Compared to HC, voice-hearers showed higher activity of the bilateral middle temporal gyrus (MTG), left fusiform gyrus and middle frontal gyrus (MFG), right superior parietal gyrus and uncus, while activity of the bilateral superior frontal gyrus (SFG) and right caudate was decreased while listening to laughing. Finally, when considering cortical responses while listening to crying, voice-hearers showed decreased activity of the hippocampii, left amygdala, IFG and insula compared to trait-negative Sz, and increased activity of bilateral SFG, and decreased bilateral amygdala, left MFG and right putamen, compared to HC.

3.4.2.1.4. Clinical studies of passively listening to affective single words with AVH-mimicking content. Escartí et al. (2010) examined the fMRI data of 27 Sz voice-hearers, 14 individuals with trait-negative Sz and 31 HC using an independent component analysis of whole brain images. Cortical activity was contrasted between passively listening to neutral words and highly affective words of varied valence that mimicked the typical AVH content of voice-hearers in the study, which were spoken in a neutral and emotional tone respectively. As compared to neutral words, listening to the AVH mimics resulted in activity of several functional networks; in voice-hearers, these were temporal, fronto-parietal, fronto-temporal, limbic, and occipito-cerebellar networks. Compared to both non-AVH groups, hyperactivity of the limbic network was observed in the voice-hearers. Overall, increased activity of the amygdala and parahippocampus gyrus, decreased activity of the STG and IFG, and a loss of activation in the left insula was observed in voice-hearers. Extending on this study, de la Iglesia-Vaya et al. (2014) investigated synchrony and effective connectivity of these networks *via* timewise correlations of these data. The researchers found differential patterns of cortico-cortical functional synchrony in the voice-hearers compared to both non-AVH groups. This was seen as pronounced delays in the activity of: (i) the occipito-cerebellar and limbic networks, which were significantly correlated, and; (ii) the fronto-temporal, temporal and temporo-parietal networks, which were significantly correlated. With a Granger causality analysis, the dominant causal source of network synchrony in the voice-hearing group was effective connectivity of the occipito-cerebellar network, whereas effective connectivity within the temporal network drove activity in both non-AVH groups.

The effects of neutral and AVH-mimicking words with varied connotation and valence were investigated by Sanjuan et al. (2007) during a study of 11 Sz voice-hearers and 10 HC. As above, AVH-mimicking words were spoken in an emotional tone. During a whole brain analysis with fMRI, passively listening to AVH mimicking speech, compared to neutral speech, elicited higher activity in several cortical regions in voice-hearers. This was most notable in the left MTG, but also occurred in the bilateral insula, orbital middle frontal cortex and PCC, right amygdala, median cingulate, orbital IFG, STG and superior medial cortex. Neutral words, compared to silence, resulted in higher activity of the left middle cingulate, MTG, orbital IFG and STG in voice-hearers after liberal thresholding (no correction for multiple comparisons). Comparatively, HC showed no differences in cortical activity between the two conditions. With the same task as above and an expanded sample, Martí-Bonmatí et al. (2007) employed a coincidental analysis of whole brain fMRI and structural MRI in 21 Sz voice-hearers and 10 HC. While listening to the AVH-mimicking speech, voice-hearers showed greater activity of the bilateral MTG, left middle cingulum, right A1, angular gyrus, PCC, STG, superomedial frontal gyrus and thalamus than HC. Structurally, lower cortical density was observed in the insula and lingual gyrus bilaterally, left MTG and postcentral gyrus, right precuneus and superomedial frontal gyrus of voice-hearers. Coincidental analysis of the functional and structural maps revealed larger clusters of activation in the bilateral MTG and STG, and smaller clusters of activation in the left inferior opercularis and PCC and right ACC and middle occipital gyrus than HC.

3.4.2.1.5. Clinical studies of actively listening to affective single words. Using an overlapping sample of 10 Sz voice-hearers, 10 individuals trait-negative Sz and 11 HC, Allen et al. (2007) and Mechelli et al. (2007) assessed source discrimination with fMRI. For this source discrimination task, emotional and neutral adjectives describing people were pre-recorded from participants and presented in four conditions: in self- or alien-speech, with or without pitch distortion. The emotional valence of words was described as either positive or negative, and each condition had an even distribution of word pitch, frequency and valence. In both studies, voice-hearers showed higher attribution errors for undistorted self-voice speech regardless of affectivity. Using a whole brain analysis, Allen et al. (2007) reported that voice-hearers showed decreased left STG activity when processing alien voices compared to both non-AVH groups. As a result of this decrease, similar levels of activity were recorded in the left STG of voice-hearers regardless of speech source. Further differences were noted in the cingulate gyrus while processing distorted speech; cingulate activity in response to distorted and undistorted speech was similar in voice-hearers, whereas both non-AVH groups showed higher activity while processing distorted speech. When voice-hearers listened to their own distorted voice activity of the right STG was increased and when alien voices were distorted, a loss of right ACC activity was observed. A positive correlation between left MTG activity and accuracy on the attribution task was absent in voice-hearers, which was observed in both non-AVH groups. Extending on these results, and including the data of one additional voice-hearing participant, Mechelli et al. (2007) implemented dynamic causal modelling to investigate intrinsic and functional connectivity of the ACC, IFG and STG. The only AVH-specific

Table 4
Recommendations for improving the quality of research into external speech processing in voice-hearers.

Category	Recommendations
Participant samples	
Group status	Clearly defined and reported AVH groups. For example, participant groups should be classified based on: state-positive (currently hallucinating); trait-positive (a history of AVH but not actively hallucinating) and; those that have never hallucinated but have a clinical history (at most, individuals had a SAPS rating of < 3 for a maximum of one week during a psychotic episode). Ideally, both trait-positive and never hallucinators should be included in separate groups alongside voice-hearers in future studies.[a]
Standardised measures of AVH severity	It is recommended that AVH-specific sub-scores of standardised clinical assessments such as PSYRATS or SAPS are used future studies. Gross measures of positive symptomology (*e.g.* PANSS overall positive score) are not specific enough to define AVH severity.
Scoring transparency	The AVH scoring criteria (*e.g.* cut-off scores on a standardised clinical assessment to confirm the presence of AVH) and the group-averaged results of these assessments should be published alongside other demographic information.
Different voice-hearing groups	We suggest that more comparisons begin to be made into voice-hearers transdiagnostically, considering: (i) comparisons between clinical and non-clinical voice-hearers and, (ii) different clinical voice-hearing cohorts, including neurodegenerative diseases and mood disorders.[b]
Power	It is recommended that future studies use *a priori* power analyses or, alternatively, pertinent publications in the field to determine suitably powered sample sizes. Appropriate statistical techniques (*e.g.* Bayesian modelling, controlling the false discovery rates) may also assist in overcoming small, underpowered samples.[c]
Group characteristics relative to listening	Testing the hearing range of participants prior to testing will eliminate any differences in responses to tasks caused by hearing loss.
Speech processing tasks	
Accessibility & open science principles	It is recommended that any study-specific tasks used are made freely available (if permissible) to enable study replication and decrease the heterogenous methodologies seen in the field.[c]
Reporting of external speech task	As much characterising detail as possible to typify the speech presented in listening tasks should be reported in the methodology of future studies. As a minimum, it is suggested that future studies report the dB[d] and gender[e] of speech presentation. Publishing examples of words or sentences used in these tasks would also enable study replication.[d]
Using a range of tasks with increasing linguistic complexity or cognitive load	For example, external speech tasks which investigate speech comprehension, contrasts between emotional and AVH mimicking language, differing emotional valences (*e.g.* positive *versus* negative), long strings of sentences (*e.g.* stories, conversations), memory processes, metaphor, semantics. Comparisons between the outcomes of these tasks, and between these tasks and passive listening, will improve our quality of understanding.
Noise cancelling	We recommend using noise-cancelling or air-gapped headphones while presenting speech to participants to minimise scanner noise.[f]
Neuroimaging techniques	
Multimodal neuroimaging	Comparisons between temporally and spatially accurate data, or between functional and structural correlates where seed-based or ROI approaches are used.
Whole brain analyses	Including the cerebellum; data-driven approaches of network-level interactions with whole brain data.

Note: AVH: auditory verbal hallucination; dB: decibels; PANSS P: positive sub-scores of the Positive and Negative Syndrome Scale; PSYRATS: Psychotic Symptom Rating Scale; ROI: region of interest; SAPS: Scale for the Assessment of Positive Symptoms.

[a] Based on findings in Alderson-Day et al. (2015), Kühn and Gallinat (2012), Toh et al. (2020a), Woodruff et al. (1997).

[b] Based on findings in Allen et al. (2008), Moseley et al. (2020b), Rolland et al. (2014).

[c] Based on recommendations in Poldrack et al. (2017).

[d] Volumes higher or lower than a normal speaking volume (70−75 dB) can lead to lower accuracy on cognitive tasks (Lausen and Hammerschmidt, 2020) and are often used to convey emotion (Kamiloğlu et al., 2020; Scherer, 2018). Evidence suggests that the primary auditory cortex has a loudness-dependent response (Mulert et al., 2005).

[e] Based on recommendations in (Kamiloğlu et al., 2020). Speaker gender can influence interpretation of affective content (Kamiloğlu et al., 2020; Lausen and Schacht, 2018) and some evidence suggests that the average frequency of different voice genders results in differential activation of the primary auditory cortex (Doucet et al., 2019).

[f] Based on comments in Kompus and Hugdahl (2018): scanner noise may present a confound as different groups may habituate to the repetitive noise at different rates.

differences observed in this study were when considering speech source. When listening to self-speech, functional connectivity between the left ACC and STG was increased in voice-hearers, whereas this connection was stronger when listening to alien speech in both non-AVH groups. Compared to neutral words, no effects of emotional language were observed in either study.

Pinheiro et al. (2017) investigated self and alien speech with a positive, negative or neutral valence in 15 SSD voice-hearers and 16 HC using EEG. All words were adjectives and spoken in a neutral tone. Voice-hearers exhibited poorer accuracy in discriminating the source of negative speech, while no differences were reported for positive or neutral speech. N100, P2 and late positive (LP; 500–700 ms post-stimulus) auditory-evoked ERP amplitudes were analysed at frontocentral, central and centroparietal electrode clusters. Overall LP amplitudes were decreased when listening to positive or neutral self-speech, and all negative speech resulted in lower centroparietal LP amplitudes in voice hearers. AVH severity, as per 'voice conversing' scores on SAPS predicted a larger difference in negative *versus* positive LP amplitude, whereas PANSS hallucination severity predicted larger LP amplitudes when listening to negative alien speech. Two effects of P2 amplitudes were found in voice-hearers: a loss of P2 increases were observed when listening to self, compared to alien, speech when the content was positive and; larger frontocentral P2 amplitudes were observed when listening to neutral alien speech than neutral self-speech, whereas no differences were reported in HC. No differences in hemispheric laterality or N100 were observed between the two groups or across electrode clusters.

3.4.2.2. Non-clinical studies of processing single words

3.4.2.2.1. Non-clinical studies of actively listening to single words. During fMRI scanning, the effects of passive listening and bimodal selective attention were examined in 12 hypnogogic/hypnopompic non-clinical voice-hearers and 12 HC by Lewis-Hanna et al. (2011). During the attention task, mismatched visual and auditory numbers were presented, with participants instructed to attend to either the auditory or visual cue, whereas the passive listening task presented short phrases to participants. With a voxel of interest (VOI) analysis restricted to the ACC, selective auditory attention elicited higher activity of right ACC in voice-hearers than HC. With a VOI analysis restricted to the middle, superior and transverse temporal gyri and the posterior temporoparietal cortex (TPC), during passive listening, voice hearers showed heightened activity of the left TPC, most significantly in the supramarginal gyrus (SMG).

3.4.3. Strings of words and sentences

Sixteen studies investigated cortical function using meaningful strings of words such as definitions, commanding sentences and stories (Table 3). Twelve studies investigated voice-hearers with Sz, three investigated voice-hearers during the first or second psychotic episode and one final study investigated non-clinical voice-hearers. Among these were seven studies of passive listening, four with active listening, four with affective, AVH-mimicking content and one with both active listening and affective, AVH-mimicking content. Thirteen studies investigated cortical function with fMRI, one with EEG, one with MEG and one with PET. Deviations in activity within the auditory, language processing and salience networks, particularly in temporal regions, were observed in many studies. Hyperactivity of limbic regions was also reported with the inclusion of affective content.

3.4.3.1. Clinical studies of processing strings of words

3.4.3.1.1. Clinical studies of passively listening to strings of words. Woodruff et al. (1997) used a story to assess passive listening in four different groups: (i) seven Sz voice-hearers, who were compared during a period of acute AVH symptomology (state-positive) and a period of mild AVH symptomology proceeding this (state-negative); (ii) eight individuals with Sz with a history of AVHs who were not actively hearing-voices (trait-positive); (iii) seven individuals with trait-negative Sz and; (iv) eight HC. With a partial brain analysis of fMRI data focussed on frontal, temporal and occipital regions, Woodruff et al. (1997) found decreased activity of the right MTG and left STG during passive listening in periods of acute AVH symptomology compared to participants' state-negative contrasts. In both trait-positive voice-hearers and trait-negative Sz individuals compared to HC, decreased activation of the left associative auditory cortex and STG and increased activation of the right MTG were observed. Also implementing passive listening with a story, Briend et al. (2017) assessed the cortical activity of 11 Sz voice-hearers and 10 HC, probing functional connectivity of the A1 and temporal gyrus with a seed-based analysis of fMRI data. Voice-hearers showed decreased functional connectivity between the A1 homologues, bilateral temporal gyri, and between right A1 and right temporal gyrus during the task compared to HC, which persisted after 20 Hz repetitive transcranial magnetic stimulation over the left superior temporal sulcus (STS).

Comparatively, Jardri et al. (2011a) used a poem to investigate the effects of self, alien and reversed (*i.e.* unintelligible) speech in 15 Sz voice-hearers and 15 HC with fMRI. During the 'self' condition, participants were instructed to mentally repeat the poem, whereas during unintelligible and alien conditions, passive listening was used. When listening to all intelligible language, the authors identified a network characterised by activity of bilateral Broca's area and middle precentral gyrus, left insula, MTG and SMA and right inferior parietal lobe (IPL) in both groups. After controlling for education, voice-hearers showed heightened activity in the right hemisphere of this network. Right IPL positively correlated with PANSS positive scores in voice-hearers, which remained significant after controlling for chlorpromazine equivalence. Additionally, increased ACC and PCC, right medial parietal gyrus, MTG, IPL and SMA activity, alongside smaller voxel clusters in the right medial parietal gyrus, MTG and IPL, were recorded in voice-hearers compared to HC.

The remainder of studies used single sentences during their investigations. Monoaural exposure to commanding statements spoken by familiar and unfamiliar voices was assessed by Zhang et al. (2008b) in 13 voice-hearers and 13 trait-negative individuals during their first or second psychotic episode and 13 HC. All commanding statements referred to movement, audition or visualisation (*e.g.* "open the curtains"; "listen to the discussion" or; "look at the van"). A whole brain analysis of fMRI data revealed an overall increase in activity in left angular gyrus, SMG, STG and STS in voice-hearers compared to either non-AVH group, which did not differ. Additionally, during right ear presentation, voice-hearers displayed a loss of activity in right MFG compared to HC.

The final three studies investigated passive listening in conjunction with paired visual stimuli. While looking at images of commonly occurring objects, participants passively listened to a corresponding definition (*e.g.*, pillow: "something you rest your head on when sleeping"). These studies assessed network-level activity with a constrained principle component analysis of fMRI data. Lavigne et al. (2015) compared 10 voice-hearers with Sz to 13 state-negative voice hearers with Sz, 27 HC and 22 bipolar disorder participants who did not hear voices. During the task, two functional networks were isolated: (i) a language processing network, consisting of increased activity of the left pars opercularis of the IFG, the bilateral STG, fusiform gyrus, visual cortex and supplementary motor area (SMA) and; (ii) a network including components of the default mode network, consisting of increased bilateral fusiform gyrus and visual cortex activity, and decreased inferior parietal cortex (IPC), lateral occipital cortex, PCC, medial PFC, precentral gyrus, precuneus, SFC, and superior parietal cortex (SPC) activity. Voice-hearers showed increased activity of these networks between 0–7.5 and 20–22.5 s compared to all other groups, between 0–5 and 15–22.5 s compared to Sz state-negative voice-hearers, and between 0–7.5 and 15–20 s compared to individuals with bipolar disorder. A positive correlation between SSPI hallucination score and cortical activity was also reported for both networks.

Following this, Lavigne and Woodward (2018) split the above Sz sample into 12 voice-hearers and 11 state-negative voice-hearers, alongside the same 27 HC. A different cut-off value of hallucination scores on the Signs and Symptoms of Psychotic Illness (SSPI) rating scale was used to categorise Sz participants as voice-hearers or state-negative voice-hearers. Slight differences in the isolated functional networks were reported. These were: (i) an auditory-motor network, characterised by increased activity of the bilateral cerebellum, insula, STG, SMA, temporal pole, thalamus and visual cortex, dorsal ACC and left precentral gyrus; (ii) a language processing network, characterised by increased activity of the bilateral visual cortex, left dlPFC, IFG, orbitofrontal cortex and posterior MTG and; (iii) a network including components of the default mode network, characterised by increased activity of bilateral A1, SMA, STG and visual cortex and left precentral gyrus, and decreased activity of the bilateral lateral occipital cortex, PCC, precuneus, ventromedial prefrontal cortex (vmPFC). Voice-hearers showed increased activity of these networks compared to both non-AVH groups at 6.25 and 8.75 s and compared to HC at 3.75 s. The Sz state-negative voice-hearers showed higher activity at 13.75 s compared to both other groups.

In the final study using a similar task, Rapin et al. (2012) compared five voice-hearers with SSD to 10 HC. After passive listening, one functional network significantly differed between the groups. This was comprised of increased activity most notably in the bilateral STG, left planum temporale and right MTG, but also in the bilateral fusiform gyrus, lingual gyrus, intracalcarine cortex and occipital pole, alongside decreased activity of the bilateral angular gyrus and MFG, left frontal pole, insula, pars opercularis and pars triangularis of the IFG, SFG, superior occipital gyrus and thalamus and right inferior occipital gyrus, SMG, superior parietal lobe and cerebellar regions. Voice-hearers showed increased activity of this network between 10–12.5 s compared to HC.

3.4.3.1.2. Clinical studies of actively listening to strings of words or sentences. Both Zhang et al. (2008a) and Mou et al. (2013) investigated the same groups of HC, voice-hearing and trait-negative individuals during their first or second psychotic episode as Zhang et al. (2008b; see Section 3.4.3.1.1). With the same commanding statements spoken by familiar or unfamiliar voices, using a source discrimination task, Zhang et al. (2008a) found that voice-hearers displayed higher discrimination errors than both non-AVH groups, mistaking familiar speech as alien. With fMRI, voice-hearers showed decreased right STG activity when listening to familiar speech compared to HC. No group differences in cortical activity were reported for alien speech. Mou et al. (2013)

performed a seed-based functional connectivity analysis of the right STG during the same task. Compared to both non-AVH groups, voice-hearers showed decreased connectivity between right STG and right SFG, and this was positively correlated with discrimination accuracy. Compared to HC, decreased connectivity was also observed between right STG and left MFG in voice-hearers.

Using a source discrimination task, Stephane et al. (2018) investigated the differences in cortical activity between self- and alien-generated sentences in seven Sz voice-hearers and eight HC. During scanning, participants were asked to discriminate which sentences they had or had not spoken during an earlier testing phase. The authors confirmed the presence of voice-hearing *via* personal communication; however, no formalised measures were reported, and it was not clear if a partial or whole brain fMRI analysis was used. Accuracy on the task was also not reported. Voice-hearers showed increased activity of the ACC, insula and SMA while listening to self-speech compared to alien speech, whereas the opposite trend was recorded in HC.

Investigating 15 voice-hearers with Sz, Plaze et al. (2006) used a recall task to ensure participants were actively listening to speech presented during an fMRI scan. After passively listening to sentences, participants indicated if phrases had been a part of the preceding sentence. Using a whole brain analysis, the researchers found that compared to silence, actively listening to speech resulted in heightened activity of the bilateral STS and IFG. Decreased activity of the left STG negatively covaried with scores on the Psychotic Symptom Rating Scale (PSYRATS), and also with auditory hallucination scores on the SAPS after small volume corrections.

3.4.3.1.3. Clinical studies of passively listening to affective sentences with AVH-mimicking content. Ford et al. (2002) examined a group of seven Sz voice-hearers, five Sz non-voice-hearers and 10 HC while listening to AVH-mimicking statements during an EEG recording. It was unclear if non-voice-hearing individuals were state- or trait-negative for AVH. Participants both passively listened to pre-recorded statements of themselves repeatedly speaking the AVH mimics or spoke the statements aloud. All AVH mimics were spoken in a neutral tone and varied in connotation and valence (*e.g.* "this is going to work out perfectly"; "why are you trying to annoy me?"). Coherence of delta, theta, alpha, beta and gamma bands were assessed between frontal and temporal electrode sites, alongside theta band power. No increases in any frequency band were observed during listening compared to talking. During talking, compared to listening, an AVH-specific reduction in theta coherence between lateral frontal and posterior temporal sites was recorded compared to both non-AVH groups.

During a MEG scan, Haesebaert et al. (2013) investigated negative AVH mimicking statements similar to those described above in Ford et al. (2002), in six Sz voice-hearers and 12 HC. Four conditions were used: passive listening, vocal repetition, inner speech repetition and silence, to assess N100 amplitudes from temporal regions, which were compared to white noise and a neutral speech sound. During both listening to and inner speech of AVH mimicking statements, left temporal M100 was reduced in voice-hearers compared to HC.

Horga et al. (2014a) assessed a group of nine individuals with Sz who previously experienced AVH during an acute phase of their illness (trait-positive, state-negative) and eight HC. All AVH mimicking statements were individually matched for the perceived content, gender and tone of AVH reported by voice-hearing participants, with each HC participant listening to phrases mimicking the AVH of a voice-hearer in the study. AVH mimics were predominantly critical or derogatory comments, ranging from single words to complex discourse. This study used ^{18}F-FDG PET to assess relative glucose metabolism rates during passive listening. Compared to HC, voice-hearers exhibited increased glucose metabolism of the bilateral hippocampus, thalamus, amygdala, left orbitofrontal cortex, right STG, brainstem and cerebellar vermis, alongside decreased glucose metabolism of the fusiform gyrus. Increased activity of the bilateral amygdala was also associated with increased activity of the bilateral A1 in voice-hearers. Additionally, activity of the left amygdala was associated with increased activity of the hippocampus, medial geniculate nucleus and posterior thalamus, and decreased activity of the bilateral medial PFC and precuneus and right MTG in the voice-hearing group compared to HC.

3.4.3.1.4. Clinical studies of actively listening to affective sentences with AVH-mimicking content. Comparing 10 voice-hearers with SSD to 10 HC, Horga et al. (2014b) assessed whole brain cortical activity with fMRI. AVH-mimicking speech included derogatory or neutral statements of varied linguistic complexity and clarity, spoken by one or multiple speakers. To ensure active listening, a low-demand task required participants to confirm the presence of AVH mimicking speech among non-speech stimuli. This study fit predictive coding models to the fMRI data to establish speech prediction signals and deficits in these signals. Increased speech prediction signalling in right A1 was observed in voice-hearers compared to HC during silence. While listening to the AVH-mimicking speech, voice-hearers showed poor prediction error signalling in the right STS and MTG, close to the right A1, compared to HC. The severity of AVH, as per PSYRATS AVH scores, was positively correlated with both higher activity during silence and magnitude of predictive error deficits in right A1. Experiencing AVH during scanning, compared to silence, was associated with lower activity of the posterior thalamus and ventral tegmental area, alongside higher activity of the left STS after small volume correction.

3.4.3.2. Non-clinical studies of processing strings of words

3.4.3.2.1. Non-clinical studies of actively listening to strings of words or sentences. Alderson-Day et al. (2017) used distorted sentences, both partially intelligible and unintelligible sine wave speech, during an fMRI scan of 12 non-clinical voice-hearers and 17 HC. Participants were instructed to listen for target sounds amongst noise and were not informed of the presence of speech. With reference to visual markers, following the scan, voice-hearers reported recognising intelligible speech significantly faster than HC, which positively correlated to 'physical characteristics' PSYRATS scores. Increased speed of speech recognition was also associated with higher cortical activity in the rostral ACC, extending into the bilateral middle cingulate and pre-SMA and left SFG. A follow-up Bayesian analysis restricted to A1 revealed no differences in activity between the voice-hearers and HC.

4. Discussion

This review aimed to systematically synthesize the neurobiological correlates of external speech processing deficits in individuals who hear voices. By approximating studies across the hierarchy of language, and with an added advantage of analysing both spatially and temporally sensitive data, some noteworthy trends were uncovered. Atypical functional activity of temporal auditory regions was reliably detected amongst voice-hearers whilst processing external speech. These deviations were observed while both passively listening to and actively processing speech sounds, words and sentences, suggesting they span the language hierarchy and are an intrinsic component of the neural alterations seen in AVH. While some contradictory findings were observed, decreased left STG activity, aberrant A1 activity and imbalanced interhemispheric coupling of A1, proportionate to AVH severity, appear to underlie external speech processing deficits in voice-hearers, as depicted in Fig. 2. However, considerable methodological variability and poor study quality limited the validity and generalisability of many reviewed publications. Here, clinical voice-hearing studies are considered first before the non-clinical literature is integrated.

Studies examining passive listening in clinical voice-hearers with speech sounds, single words and sentences implicated similar temporal auditory regions, although the directionality of changes in activity were sometimes contested. For example, during passive listening to speech sounds, increased gamma coupling of A1 homologues was detected amongst Sz voice-hearers *via* a source estimation technique (Steinmann et al., 2017); which the authors argued to be a marker of increased functional connectivity (Tagliazucchi et al., 2012). However, decreased A1 functional connectivity has also been detected amongst voice-hearing patients with Sz whilst passively listening to words and sentences (Briend et al., 2017; Gavrilescu et al., 2010). Other studies found AVH-specific differences only in the right A1 homologue. Decreased right A1 activity during listening (Martí-Bonmatí et al., 2007), and increased right A1 activity, proportionate to the severity of AVH, during silence (Horga et al., 2014b) have also been detected amongst Sz voice-hearers. Right ear presentation of words was also related to suppression of N100 amplitudes over right temporal regions in voice-hearers (Innes-Brown et al., 2006). Taken together, the reviewed literature implicates bilateral A1 involvement in the dysfunctional processing of external speech amongst clinical voice-hearing individuals.

Extending on A1, wider involvement of the STG in speech processing deficits appear to span the language hierarchy in clinical voice-hearers. A decrease in the healthy left-sided lateralisation of the STG during dichotic listening in voice-hearers with temporal lobe epilepsy (Korsnes et al., 2010) was also supported by findings of a loss of right ear advantage in Sz voice-hearers during the same task (*e.g.* Steinmann et al., 2017); a reliable marker of a loss of healthy left-favouring STG asymmetry or poor interhemispheric communication (Hugdahl et al., 2008; Steinmann et al., 2014a). In agreeance with findings from speech sound studies, decreased left STG activity in Sz voice-hearers, proportionate to AVH severity, was evident during listening to sentences, regardless of active or passive listening (Plaze et al., 2006; Woodruff et al., 1997). EEG studies showed decreased coupling of temporal regions in Sz voice-hearers, seen through both increased interhemispheric transfer time of the N100, and decreased coherence of the upper alpha band (Henshall et al., 2012, 2013). Studies investigating network-level interactions with data-driven approaches in psychotic voice-hearing groups also suggest that increased coupling of the A1 and STG to a wider network, including components of the auditory, language processing and default mode networks, underlie AVH-specific speech processing deficits with passive listening (Lavigne et al., 2015; Lavigne and Woodward, 2018), also appearing in active listening tasks (Mechelli et al., 2007; Mou et al., 2013). In support of this, decreased coherence of the upper beta band and increased late positive potentials in Sz voice-hearers were argued by the authors to suggest impairment of the coordination of auditory information and aberrant attentional processes respectively (Henshall et al., 2013; Pinheiro et al., 2017).

When including active listening in tasks with speech sounds, words or sentences, several further changes in cortical activity were observed in clinical voice-hearers. Actively listening to speech sounds resulted in a suppression of N100 amplitude in the left hemisphere, which was proportionate to the severity of AVH (Heinks-Maldonado et al., 2007). There is no evidence to suggest that clinical voice-hearers experience difficulty in identifying the gender of a speaker (Kang et al., 2009). However, decreased abilities for source recognition, including self-speech (*e.g.* Allen et al., 2007; Zhang et al., 2008a), may be further impacted by distortion in psychotic voice-hearing groups (*e.g.* Heinks-Maldonado et al., 2007). Decreased functional connectivity between the right STG and right SFG appear involved in poor recognition abilities in psychotic voice-hearers (Mou et al., 2013), which may be caused by deviant activity of the right temporal regions (Zhang et al., 2008a). Investigations of AVH-specific cognitive abilities during memory tasks are underrepresented in the literature, making it challenging to draw related conclusions. However, although voice-hearers showed no differences in accuracy, some evidence suggests that Sz voice-hearers show hyperactivity of the posterior basal ganglia during simple recall tasks (Ikuta et al., 2015).

While processing emotional or AVH-mimicking speech, Sz voice-hearers appear to exhibit deviant functioning in the primary language network and limbic regions. Alongside the temporal auditory processing regions, affective structures including the amygdala, hippocampus and parahippocampal gyrus appear to be involved in aberrant affective speech processing (*e.g.* Escartí et al., 2010; Horga et al., 2014a,b). However, both hyper- and hypoactive responses of these regions were recorded during affective listening tasks (Escartí et al., 2010; Kang et al., 2009). Some evidence suggests that negative emotional valence of speech produces more deviations in healthy responses than positive speech (*e.g.* Kang et al., 2009; Pinheiro et al., 2017), which may be underscored by delayed occipito-cerebellar responses in Sz voice-hearers (de la Iglesia-Vaya et al., 2014). However, it should be noted that some studies found no effect of emotional speech whatsoever (Allen et al., 2007; Mechelli et al., 2007). While some evidence suggests that right-sided lateralisation changes during affective speech processing exist in clinical voice-hearers (Allen et al., 2007; Kang et al., 2009), it is noteworthy that this was not a robustly replicated finding (*e.g.*, Escartí et al., 2010; Mechelli et al., 2007). Further research is needed to delineate the involvement of cortical regions in affective speech processing in clinical voice-hearers, in particular whether there are any laterality differences during these tasks.

Comparatively, a dearth of information exists on external speech processing in non-clinical voice-hearing populations. Alongside a considerably smaller number of studies, several key behavioural tasks remain unresearched in these individuals. For instance, none of the reviewed literature investigated cortical responses to listening tasks involving working memory or emotional speech in non-clinical voice-hearers. Furthermore, only one study of non-clinical voice-hearers with fMRI implemented a whole brain analysis (Alderson-Day et al., 2017). Nonetheless, some similarities between clinical and non-clinical voice-hearers were observed; predominantly in the temporal cortex.

Similar to the decreased right ear advantage seen in clinical voice-hearers during passive listening, decreased right ear acuity was reported in non-clinical voice-hearers in response to speech sounds (Kompus et al., 2013); again indicating a loss of left-favouring STG asymmetry or aberrant interhemispheric connectivity (Hugdahl et al., 2008; Steinmann et al., 2014a). Also similar to clinical voice-hearers, studies of speech sounds showed decreased right A1 activity (Kompus et al., 2013) and increased gamma band synchrony during left-ear presentation (Thiebes et al., 2018). Some evidence suggests that the morphology of right A1 may be altered in non-clinical voice-hearers, where a more lateral spread of voxels was activated in response to speech sounds (Kompus et al., 2013). The sole remaining finding from passive listening studies was hyperactivity of the left TPC, notably in the SMG (Lewis-Hanna et al., 2011). Hyperactive left SMG function has also been recorded in psychotic voice-hearing cohorts while passively listening to both single words and sentences (Rapin et al., 2012; Zhang et al., 2008b).

Although the study designs of each investigation of active listening in non-clinical voice-hearers were heterogenous, three notable trends arose from the literature. Firstly, contrary to Sz voice-hearers (Heinks-Maldonado et al., 2007), increased N100 responses to self-speech sounds were seen in non-clinical voice-hearers relative to healthy controls (Pinheiro et al., 2018). However, in both cases, the magnitude of deviant N100 responses were proportionate to the severity of AVH (Heinks-Maldonado et al., 2007; Pinheiro et al., 2018). Secondly, unlike the Sz voice-hearing literature, evidence suggests that while deviant A1 activity is observed in non-clinical voice-hearers during passive listening, healthy responses may remain intact during active listening (Alderson-Day et al., 2017; Kompus et al., 2013). Finally, there is no evidence to suggest that voice-hearers experience difficulty in discriminating words or sentences from noise; on the contrary, evidence suggests that non-clinical voice-hearers exhibit increased auditory acuity during relevant listening tasks, which was correlated with hyperactivity of medial regions such as the ACC (Alderson-Day et al., 2017; Lewis-Hanna et al., 2011). However, this should be interpreted with caution; some studies have demonstrated that attentional factors such as expectation or prior knowledge lead voice-hearers to hear speech which is not present among noise (Daalman et al., 2012).

Many pertinent reviews in similar fields have found AVH-specific differences in activity of the left temporal auditory regions. In line with our review, decreased left STG activity has been implicated as a trait-based marker of AVH in a meta-analysis of cortical activity in voice-hearers with Sz during both external and internal (*e.g.* inner monologue) speech tasks (Kühn and Gallinat, 2012). In contrast, during hallucinated speech, increased activity of the left STG has been found during both a systematic review of transdiagnostic voice-hearers (Allen et al., 2008) and a meta-analysis of symptom capture in Sz (Jardri et al., 2011b). To address this disparity in activation in response to different types of speech, another meta-analysis has investigated A1 function in Sz voice-hearers in response to either silence or auditory stimuli (Kompus et al., 2011). The authors concluded that left A1 is hypoactive in response to auditory stimuli and hyperactive in the resting state, when Sz voice-hearers were actively experiencing AVH. Taken with the results of our review, these bodies of work suggest that the left temporal auditory regions are differentially activated during hallucinated speech and external speech, which may reflect an attentional bias towards internally-generated auditory information in voice-hearers. This may explain the few papers in our review where increased left STG activity was reported (*e.g.* Sanjuan et al., 2007): these participants may have been actively hallucinating during imaging.

Many pertinent reviews have also highlighted the importance of connectivity of A1 or the STG in AVH. Among these, several theoretical reviews have concluded that the direction of activity changes in the left or right STG in the phenomenon of AVH in Sz may not be as important as changes to an overall, optimal level of bilateral STG connectivity (Steinmann et al., 2014b), which may be representative of imbalanced excitatory and inhibitory processes (Jardri et al., 2016; Steinmann et al., 2019). Extending on this, another noteworthy theoretical review investigating the functional connectivity of language and memory networks in voice-hearers discussed the interhemispheric coupling of A1: hyperconnectivity was observed in non-clinical and first-episode psychosis samples, whereas decreased coupling arose with Sz chronicity (Ćurčić-Blake et al., 2017). Together, and in line with our own review, these bodies of work illustrate the significance of the shift in bilateral coupling of the temporal auditory regions in the experience of voice-hearing. The same review also demonstrated that increased connectivity along the auditory and language processing networks is an integral part of active AVH symptomology (Ćurčić-Blake et al., 2017). Indeed, numerous pertinent reviews have investigated connectivity of the STG or A1 to wider networks in AVH samples. For example, hyperconnectivity of the left STG to other regions of the language processing network was found in a systematic review of resting-state functional connectivity in voice-hearers (Alderson-Day et al., 2015). The involvement of A1 and STG, among other regions of the auditory and language processing networks, and in conjunction with deviant default mode network activity, have also been highlighted in several recent papers which suggest that these networks constitute a single 'AVH network' (Geng et al., 2020; Scheinost et al., 2019; van Lutterveld et al., 2014). Although study designs limited conclusions in our review, the temporal primary auditory processing regions appear to be key to the speech processing deficits underlying AVH, with several others suggesting wider involvement of the cerebellar, language, default mode and limbic networks. More work aiming to delineate the state- and trait- differences in voice-hearing will be vital in confirming the influence of these networks in the speech processing deficits associated with voice-hearing.

A wealth of additional cortical regions and connectivity or coherence deficits were reported across the various reviewed studies. For instance, decreased right MTG (Horga et al., 2014a; Woodruff et al., 1997) and decreased left insula activity (Escartí et al., 2010; Jardri et al., 2011a; Kang et al., 2009) were evident in several studies of Sz voice-hearers. There was also evidence of hyperactivity in several subcortical structures including the medial geniculate nucleus and ventral tegmental area (Horga et al., 2014b; Ikuta et al., 2015; Lavigne and Woodward, 2018); these areas have implications in auditory processing and dopaminergic excitation (Ikai et al., 1992; Kiehl et al., 2005; Pinheiro et al., 2020). However, given the diversity of study designs, few consistent patterns were observed outside of the primary temporal auditory processing regions, namely, A1 and the STG.

With that being said, it is acknowledged that differences between clinical and non-clinical voice-hearing exist. For instance, while aberrant function of the auditory and language processing regions may be universal to voice-hearing (Alderson-Day et al., 2015; Barkus et al., 2007; Daalman and Diederen, 2013), it is likely that some aberrations in neural connectivity and function differ between clinical and non-clinical voice-hearing groups in response to external speech (Badcock and Hugdahl, 2012; de Leede-Smith and Barkus, 2013). The effects of antipsychotic medication also play a confounding role in deviant cortical function in clinical voice-hearers (Bolding et al., 2012; Lui et al., 2010; Sambataro et al., 2010). Outside of the temporal auditory regions, significantly more research is needed to delineate the similarities and differences in cortical function in different AVH cohorts.

4.1. Limitations

Several facets of study design hindered our ability to interpret results. Overall, the current field lacks suitable samples, with most studies drawing from small, poorly characterized participant pools. Most detrimental was the lack of inclusion of a comparable non-voice-hearing clinical group in many clinically-based studies. Where such a group was present, many did not succinctly define if comparison participants were trait- or state-negative for voice-hearing or included both trait- and state-negative individuals in the same group. This discrepancy presented a challenge when determining which outcomes were specific to voice-hearing or were merely the result of a diagnosis of Sz. Additionally, one-fifth (n = 8) of studies defined their AVH population based on overall positive symptomology (*e.g.* PANSS positive score) instead of AVH-specific scores on standardised clinical assessments such as SAPS or PSYRATS, and another quarter (n = 10) did not provide the results of AVH severity after performing an appropriate AVH-specific assessment. Over one-third (n = 8) of studies which included a state- or trait-negative clinical comparison reported comparable levels of AVH or positive symptomology between these individuals and the AVH cohort. Again, this element of methodology presented a significant challenge when determining the specificity of findings to voice-hearing. In light of this, the discussion above focussed on the 23 studies where either these distinctions were made or where correlations between AVH severity and cognitive or cortical findings were investigated. The studies that did not make these distinctions did not provide a consensus, often implicating numerous further, unreplicated cortical regions. It should also be noted that small sample sizes remain a significant issue in the field, with 82 % of studies (n = 32) investigating 15 participants or less per group. Additionally, little replication exists across study designs, particularly in terms of listening tasks. This may have been an underlying cause of the heterogenous findings between many studies. Furthermore, these methodological limitations appeared more regularly with tasks that included active listening or affective content, which limited related conclusions.

4.2. Future directions

To advance the field, a number of considerations are recommended during the development of future study protocols. Pertinently, this involves inter-laboratory coherence in methodological design and AVH classification. To address this, the field may benefit from a large multi-site study, similar to undertakings examining the changes to cognitive abilities associated with AVH (Moseley et al., 2020a) and tonal processing in Sz voice-hearers (Ford et al., 2009). The recommendations for improving study protocols outlined in Table 4 are in line with numerous reviews in related fields (*e.g.* Alderson-Day et al., 2015; Allen et al., 2008; Kühn and Gallinat, 2012). Alongside multimodal neuroimaging approaches, a comprehensive understanding of the speech processing capabilities across the hierarchy of language of voice-hearers would benefit our knowledge of the voice-hearing experience and the involvement of numerous cortical regions in this phenomenon. This would include developing a deeper understanding of speech processing difficulties in relation to models of voice perception (Aglieri et al., 2018; Belin et al., 2000). Research committing to these distinctions may address some of the problematic heterogeneity of Sz diagnoses (Allsopp et al., 2019) and the high rates of treatment-resistance among voice-hearers with Sz (Buckley and Miller, 2017; Shergill et al., 1998).

Overall, this review suggests that AVH-specific differences in cortical activity in response to listening to external speech span the hierarchy of language and are present in auditory and language regions. Further regions and networks are implicated with active listening or affective content. However, future research must succinctly define their study populations, investigate complex language processing tasks and implement whole brain, data-driven neuroimaging analyses. Through these improvements to study protocols, the involvement of cortical regions and networks in AVH-related speech processing deficits will be further elucidated. This may increase our understanding of the involvement of these regions and networks in the experience of voice-hearing. These methodological adaptations may additionally aid in our understanding of the real-life challenges that voice-hearers experience during spoken or social communication, which may have benefits for psychiatric intervention.

Funding

This work was supported by the Australian Government Research Training Program (S.E.R), the Australian National Health and Medical Research Council (NHMRC; fellowship to S.L.R (ID: 1154651), and a project grant to S.L.R (ID: 1060664)), a Barbara Dicker Brain Sciences Foundation project grant (S.L.R), the Australian National Imaging Facility, a National Collaborative Research Infrastructure Strategy (NCRIS) capability (M.E.H) and the Canadian Institutes for Health Research and British Columbia Schizophrenia Society (T.S.W).

.

Declaration of Competing Interest

The authors have no conflicts of interest to declare.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.neubiorev.2021.09.006.

References

Aglieri, V., Chaminade, T., Takerkart, S., Belin, P., 2018. Functional connectivity within the voice perception network and its behavioural relevance. NeuroImage 183, 356–365. https://doi.org/10.1016/j.neuroimage.2018.08.011.

Alderson-Day, B., McCarthy-Jones, S., Fernyhough, C., 2015. Hearing voices in the resting brain: a review of intrinsic functional connectivity research on auditory verbal hallucinations. Neurosci. Biobehav. Rev. 55, 78–87. https://doi.org/10.1016/j.neubiorev.2015.04.016.

Alderson-Day, B., Lima, C.F., Evans, S., Krishnan, S., Shanmugalingam, P., Fernyhough, C., Scott, S.K., 2017. Distinct processing of ambiguous speech in people with non-clinical auditory verbal hallucinations. Brain 140 (9), 2475–2489. https://doi.org/10.1093/brain/awx206.

Aleman, A., Larøi, F., 2008. Hallucinations: the Science of Idiosyncratic Perception. American Psychological Association, Washington, DC, US.

Allen, P., Amaro, E., Fu, C.H.Y., Williams, S.C.R., Brammer, M.J., Johns, L.C., McGuire, P.K., 2007. Neural correlates of the misattribution of speech in schizophrenia. Br. J. Psychiatry 190 (2), 162–169. https://doi.org/10.1192/bjp.bp.106.025700.

Allen, P., Larøi, F., McGuire, P.K., Aleman, A., 2008. The hallucinating brain: a review of structural and functional neuroimaging studies of hallucinations. Neurosci. Biobehav. Rev. 32 (1), 175–191. https://doi.org/10.1016/j.neubiorev.2007.07.012.

Allsopp, K., Read, J., Corcoran, R., Kinderman, P., 2019. Heterogeneity in psychiatric diagnostic classification. Psychiatry Res. 279, 15–22. https://doi.org/10.1016/j.psychres.2019.07.005.

Anderer, P., Semlitsch, H.V., Saletu, B., 1996. Multichannel auditory event-related brain potentials: effects of normal aging on the scalp distribution of N1, P2, N2 and P300 latencies and amplitudes. Electroencephalogr. Clin. Neurophysiol. 99 (5), 458–472. https://doi.org/10.1016/S0013-4694(96)96518-9.

Badcock, J.C., Hugdahl, K., 2012. Cognitive mechanisms of auditory verbal hallucinations in psychotic and non-psychotic groups. Neurosci. Biobehav. Rev. 36 (1), 431–438. https://doi.org/10.1016/j.neubiorev.2011.07.010.

Barkus, E., Stirling, J., Hopkins, R., McKie, S., Lewis, S., 2007. Cognitive and neural processes in non-clinical auditory hallucinations. Br. J. Psychiatry 191 (S51), s76–s81. https://doi.org/10.1192/bjp.191.51.s76.

Beavan, V., Read, J., Cartwright, C., 2011. The prevalence of voice-hearers in the general population: a literature review. J. Ment. Health 20 (3), 281–292. https://doi.org/10.3109/09638237.2011.562262.

Behrendt, R.-P., Young, C., 2004. Hallucinations in schizophrenia, sensory impairment, and brain disease: a unifying model. Behav. Brain Sci. 27 (6), 771–787. https://doi.org/10.1017/s0140525x04000184.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. Nature 403 (6767), 309–312. https://doi.org/10.1038/35002078.

Bolding, M.S., White, D.M., Hadley, J.A., Weiler, M., Holcomb, H.H., Lahti, A.C., 2012. Antipsychotic drugs alter functional connectivity between the medial frontal cortex, hippocampus, and nucleus accumbens as measured by H215O PET. Front. Psychiatry 3, 105. https://doi.org/10.3389/fpsyt.2012.00105.

Bradshaw, A.R., Bishop, D.V.M., Woodhead, Z.V.J., 2017. Methodological considerations in assessment of language lateralisation with fMRI: a systematic review. PeerJ 5, e3557. https://doi.org/10.7717/peerj.3557.

Briend, F., Leroux, E., Delcroix, N., Razafimandimby, A., Etard, O., Dollfus, S., 2017. Impact of rTMS on functional connectivity within the language network in schizophrenia patients with auditory hallucinations. Schizophr. Res. 189, 142–145. https://doi.org/10.1016/j.schres.2017.01.049.

Buckley, P.F., Miller, B.J., 2017. Personalized medicine for schizophrenia. NPJ Schizophr. 3 (2) https://doi.org/10.1038/s41537-016-0001-5.

Bühler, T., Kindler, J., Schneider, R.C., Strik, W., Dierks, T., Hubl, D., Koenig, T., 2016. Disturbances of agency and ownership in schizophrenia: an auditory verbal event related potentials study. Brain Topogr. 29 (5), 716–727. https://doi.org/10.1007/s10548-016-0495-1.

Costafreda, S.G., Brébion, G., Allen, P., McGuire, P.K., Fu, C.H.Y., 2008. Affective modulation of external misattribution bias in source monitoring in schizophrenia. Psychol. Med. 38 (6), 821–824. https://doi.org/10.1017/S0033291708003243.

Ćurčić-Blake, B., Ford, J.M., Hubl, D., Orlov, N.D., Sommer, I.E., Waters, F., et al., 2017. Interaction of language, auditory and memory brain networks in auditory verbal hallucinations. Prog. Neurobiol. 148, 1–20. https://doi.org/10.1016/j.pneurobio.2016.11.002.

Daalman, K., Diederen, K.M., 2013. A final common pathway to hearing voices: examining differences and similarities in clinical and non-clinical individuals. Psychosis 5 (3), 236–246. https://doi.org/10.1080/17522439.2013.796402.

Daalman, K., van Zandvoort, M., Bootsman, F., Boks, M., Kahn, R., Sommer, I., 2011. Auditory verbal hallucinations and cognitive functioning in healthy individuals. Schizophr. Res. 132 (2–3), 203–207. https://doi.org/10.1016/j.schres.2011.07.013.

Daalman, K., Verkooijen, S., Derks, E.M., Aleman, A., Sommer, I.E., 2012. The influence of semantic top-down processing in auditory verbal hallucinations. Schizophr. Res. 139 (1–3), 82–86. https://doi.org/10.1016/j.schres.2012.06.005.

de la Iglesia-Vaya, M., Escartí, M.J., Molina-Mateo, J., Martí-Bonmatí, L., Gadea, M., Castellanos, F.X., et al., 2014. Abnormal synchrony and effective connectivity in patients with schizophrenia and auditory hallucinations. Neuroimage Clin. 6, 171–179. https://doi.org/10.1016/j.nicl.2014.08.027.

de Leede-Smith, S., Barkus, E., 2013. A comprehensive review of auditory verbal hallucinations: lifetime prevalence, correlates and mechanisms in healthy and clinical individuals. Front. Hum. Neurosci. 7, 367. https://doi.org/10.3389/fnhum.2013.00367.

Doucet, G.E., Luber, M.J., Balchandani, P., Sommer, I.E., Frangou, S., 2019. Abnormal auditory tonotopy in patients with schizophrenia. NPJ Schizophr. 5 (1), 1–6. https://doi.org/10.1038/s41537-019-0084-x.

Egerton, A., Modinos, G., Ferrera, D., McGuire, P., 2017. Neuroimaging studies of GABA in schizophrenia: a systematic review with meta-analysis. Transl. Psychiatry 7 (6), e1147. https://doi.org/10.1038/tp.2017.124.

Escartí, M.J., de la Iglesia-Vayá, M., Martí-Bonmatí, L., Robles, M., Carbonell, J., Lull, J. J., et al., 2010. Increased amygdala and parahippocampal gyrus activation in schizophrenic patients with auditory hallucinations: an fMRI study using independent component analysis. Schizophr. Res. 117 (1), 31–41. https://doi.org/10.1016/j.schres.2009.12.028.

Ford, J.M., Mathalon, D.H., Whitfield, S., Faustman, W.O., Roth, W.T., 2002. Reduced communication between frontal and temporal lobes during talking in schizophrenia. Biol. Psychiatry 51 (6), 485–492. https://doi.org/10.1016/s0006-3223(01)01335-x.

Ford, J.M., Roach, B.J., Jorgensen, K.W., Turner, J.A., Brown, G.G., Notestine, R., et al., 2009. Tuning in to the voices: a multisite FMRI study of auditory hallucinations. Schizophr. Bull. 35 (1), 58–66. https://doi.org/10.1093/schbul/sbn140.

Ganguli, R., Carter, C., Mintun, M., Brar, J., Becker, J., Sarma, R., et al., 1997. PET brain mapping study of auditory verbal supraspan memory versus visual fixation in schizophrenia. Biol. Psychiatry 41 (1), 33–42. https://doi.org/10.1016/S0006-3223(96)00206-5.

Gavrilescu, M., Rossell, S., Stuart, S.W., Shea, T.L., Innes-Brown, H., Henshall, K., et al., 2010. Reduced connectivity of the auditory cortex in patients with auditory hallucinations: a resting state functional magnetic resonance imaging study. Psychol. Med. 40 (7), 1149. https://doi.org/10.1017/S0033291709991632.

Geng, H., Xu, P., Sommer, I.E., Luo, Y.-J., Aleman, A., Ćurčić-Blake, B., 2020. Abnormal dynamic resting-state brain network organization in auditory verbal hallucination. Brain Struct. Funct. 225 (8), 2315–2330. https://doi.org/10.1007/s00429-020-02119-1.

Haesebaert, F., Lecaignard, F., Suaud-Chagny, M., d'Amato, T., Saoud, M., Poulet, E., et al., 2013. Left auditory cortex dysfunction in hallucinating patients with schizophrenia: an MEG study. Clin. Neurophysiol. 124 (4), 823. https://doi.org/10.1016/j.clinph.2012.07.022.

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. Arch. Gen. Psychiatry 64 (3), 286–296. https://doi.org/10.1001/archpsyc.64.3.286.
Henshall, K.R., Sergejew, A.A., McKay, C.M., Rance, G., Shea, T.L., Hayden, M.J., et al., 2012. Interhemispheric transfer time in patients with auditory hallucinations: an auditory event-related potential study. Int. J. Psychophysiol. 84 (2), 130–139. https://doi.org/10.1016/j.ijpsycho.2012.01.020.
Henshall, K.R., Sergejew, A.A., Rance, G., McKay, C.M., Copolov, D.L., 2013. Interhemispheric EEG coherence is reduced in auditory cortical regions in schizophrenia patients with auditory hallucinations. Int. J. Psychophysiol. 89 (1), 63–71. https://doi.org/10.1016/j.ijpsycho.2013.05.005.
Horga, G., Fernandez-Egea, E., Mane, A., Font, M., Schatz, K.C., Falcon, C., et al., 2014a. Brain metabolism during hallucination-like auditory stimulation in schizophrenia. PLoS One 9 (1), e84987. https://doi.org/10.1371/journal.pone.0084987.
Horga, G., Schatz, K.C., Abi-Dargham, A., Peterson, B.S., 2014b. Deficits in predictive coding underlie hallucinations in schizophrenia. J. Neurosci. 34 (24), 8072–8082. https://doi.org/10.1523/JNEUROSCI.0200-14.2014.
Hugdahl, K., Sommer, I.E., 2018. Auditory verbal hallucinations in schizophrenia from a levels of explanation perspective. Schizophr. Bull. 44 (2), 234–241. https://doi.org/10.1093/schbul/sbx142.
Hugdahl, K., Løberg, E.M., Jørgensen, H.A., Lundervold, A., Lund, A., Green, M.F., Rund, B., 2008. Left hemisphere lateralisation of auditory hallucinations in schizophrenia: a dichotic listening study. Cogn. Neuropsychiatry 13 (2), 166–179. https://doi.org/10.1080/13546800801906808.
Ikai, Y., Takada, M., Shinonaga, Y., Mizuno, N., 1992. Dopaminergic and non-dopaminergic neurons in the ventral tegmental area of the rat project, respectively, to the cerebellar cortex and deep cerebellar nuclei. Neuroscience 51 (3), 719–728. https://doi.org/10.1016/0306-4522(92)90310-x.
Ikuta, T., DeRosse, P., Argyelan, M., Karlsgodt, K.H., Kingsley, P.B., Szeszko, P.R., Malhotra, A.K., 2015. Subcortical modulation in auditory processing and auditory hallucinations. Behav. Brain Res. 295, 78–81. https://doi.org/10.1016/j.bbr.2015.08.009.
Innes-Brown, H., Rossell, S., Egan, G., Copolov, D., McKay, C., Shea, T., et al., 2006. Central auditory processing deficits in patients with auditory hallucinations as shown by event-related potentials: preliminary results. Acta Neuropsychiatr. 18 (6), 293. https://doi.org/10.1017/S0924270800031343.
Jardri, R., Pins, D., Lafargue, G., Very, E., Ameller, A., Delmaire, C., Thomas, P., 2011a. Increased overlap between the brain areas involved in self-other distinction in schizophrenia. PLoS One 6 (3), e17500. https://doi.org/10.1371/journal.pone.0017500.
Jardri, R., Pouchet, A., Pins, D., Thomas, P., 2011b. Cortical activations during auditory verbal hallucinations in schizophrenia: a coordinate-based meta-analysis. Am. J. Psychiatry 168 (1), 73–81. https://doi.org/10.1176/appi.ajp.2010.09101522.
Jardri, R., Hugdahl, K., Hughes, M., Brunelin, J., Waters, F., Alderson-Day, B., et al., 2016. Are Hallucinations Due to an Imbalance Between Excitatory and Inhibitory Influences on the Brain? Schizophr. Bull. 42 (5), 1124–1134. https://doi.org/10.1093/schbul/sbw075.
Johns, L.C., Rossell, S., Frith, C., Ahmad, F., Hemsley, D., Kuipers, E., McGuire, P.K., 2001. Verbal self-monitoring and auditory verbal hallucinations in patients with schizophrenia. Psychol. Med. 31 (4), 705–715. https://doi.org/10.1017/s0033291701003774.
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. Schizophr. Bull. 40 (Suppl_4), S255–S264. https://doi.org/10.1093/schbul/sbu005.
Kamiloğlu, R.G., Fischer, A.H., Sauter, D.A., 2020. Good vibrations: a review of vocal expressions of positive emotions. Psychon. Bull. Rev. 27 (2), 237–265. https://doi.org/10.3758/s13423-019-01701-x.
Kaneta, T., 2020. PET and SPECT imaging of the brain: a review on the current status of nuclear medicine in Japan. J. Radiol. 38 (4), 343–357. https://doi.org/10.1007/s11604-019-00901-8.
Kang, J.I., Kim, J., Seok, J., Chun, J.W., Lee, S., Park, H., 2009. Abnormal brain response during the auditory emotional processing in schizophrenic patients with chronic auditory hallucinations. Schizophr. Res. 107 (1), 83–91. https://doi.org/10.1016/j.schres.2008.08.019.
Kapucu, Ö.L., Nobili, F., Varrone, A., Booij, J., Vander Borght, T., Någren, K., et al., 2009. EANM procedure guideline for brain perfusion SPECT using 99mTc-labelled radiopharmaceuticals, version 2. Eur. J. Nucl. Med. Mol. Imaging 36 (12), 2093. https://doi.org/10.1007/s00259-009-1266-y.
Keil, A., Debener, S., Gratton, G., Junghöfer, M., Kappenman, E.S., Luck, S.J., et al., 2014. Committee report: publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. Psychophysiology 51 (1), 1–21. https://doi.org/10.1111/psyp.12147.
Kiehl, K.A., Stevens, M.C., Laurens, K.R., Pearlson, G., Calhoun, V.D., Liddle, P.F., 2005. An adaptive reflexive processing model of neurocognitive function: supporting evidence from a large scale (n = 100) fMRI study of an auditory oddball task. NeuroImage 25 (3), 899–915. https://doi.org/10.1016/j.neuroimage.2004.12.035.
Kompus, K., Hugdahl, K., 2018. Perception of voices that do not exist: neuronal mechanisms in clinical and non-clinical hallucinations. In: Frühholz, S., Belin, P. (Eds.), The Oxford Handbook of Voice Perception. Oxford University Press.
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. https://doi.org/10.1016/j.neuropsychologia.2011.08.010.
Kompus, K., Falkenberg, L.E., Bless, J.J., Johnsen, E., Kroken, R.A., Kråkvik, B., et al., 2013. The role of the primary auditory cortex in the neural mechanism of auditory verbal hallucinations. Front. Hum. Neurosci. 7, 144. https://doi.org/10.3389/fnhum.2013.00144.
Korsnes, M.S., Hugdahl, K., Nygård, M., Bjørnæs, H., 2010. An fMRI study of auditory hallucinations in patients with epilepsy. Epilepsia 51 (4), 610–617. https://doi.org/10.1111/j.1528-1167.2009.02338.x.
Kühn, S., Gallinat, J., 2012. Quantitative meta-analysis on state and trait aspects of auditory verbal hallucinations in schizophrenia. Schizophr. Bull. 38 (4), 779–786. https://doi.org/10.1093/schbul/sbq152.
Larøi, F., Sommer, I.E., Blom, J.D., Fernyhough, C., Ffytche, D.H., Hugdahl, K., et al., 2012. The characteristic features of auditory verbal hallucinations in clinical and nonclinical groups: state-of-the-art overview and future directions. Schizophr. Bull. 38 (4), 724–733. https://doi.org/10.1093/schbul/sbs061.
Lausen, A., Hammerschmidt, K., 2020. Emotion recognition and confidence ratings predicted by vocal stimulus type and prosodic parameters. Hum. Soc. Sci. Commun. 7 (1), 2. https://doi.org/10.1057/s41599-020-0499-z.
Lausen, A., Schacht, A., 2018. Gender differences in the recognition of vocal emotions. Front. Psychol. 9, 882. https://doi.org/10.3389/fpsyg.2018.00882.
Lavigne, K.M., Woodward, T.S., 2018. 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. 39 (4), 1582–1595. https://doi.org/10.1002/hbm.23934.
Lavigne, K.M., Rapin, L.A., Metzak, P.D., Whitman, J.C., Jung, K., Dohen, M., et al., 2015. Left-dominant temporal-frontal hypercoupling in schizophrenia patients with hallucinations during speech perception. Schizophr. Bull. 41 (1), 259–267. https://doi.org/10.1093/schbul/sbu004.
Lewis-Hanna, L.L., Hunter, M.D., Farrow, T.F.D., Wilkinson, I.D., Woodruff, P.W.R., 2011. Enhanced cortical effects of auditory stimulation and auditory attention in healthy individuals prone to auditory hallucinations during partial wakefulness. NeuroImage 57 (3), 1154–1161. https://doi.org/10.1016/j.neuroimage.2011.04.058.
Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gøtzsche, P.C., Ioannidis, J.P.A., et al., 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Ann. Intern. Med. 151 (4), W–65. https://doi.org/10.1136/bmj.b2700.
Lui, S., Li, T., Deng, W., Jiang, L., Wu, Q., Tang, H., et al., 2010. Short-term effects of antipsychotic treatment on cerebral function in drug-naive first-episode schizophrenia revealed by "resting state" functional magnetic resonance imaging. Arch. Gen. Psychiatry 67 (8), 783–792. https://doi.org/10.1001/archgenpsychiatry.2010.84.
Martí-Bonmatí, L., Lull, J.J., García-Martí, G., Aguilar, E.J., Moratal-Pérez, D., Poyatos, C., et al., 2007. Chronic auditory hallucinations in schizophrenic patients: MR analysis of the coincidence between functional and morphologic abnormalities. Radiology 244 (2), 549–556. https://doi.org/10.1148/radiol.2442060727.
McCarthy-Jones, S., Resnick, P.J., 2014. Listening to voices: the use of phenomenology to differentiate malingered from genuine auditory verbal hallucinations. Int. J. Law Psychiatry 37 (2), 183–189. https://doi.org/10.1016/j.ijlp.2013.11.004.
Mechelli, A., Allen, P., Amaro Jr., 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. Hum. Brain Mapp. 28 (11), 1213–1222. https://doi.org/10.1002/hbm.20341.
Merrett, Z., Rossell, S.L., Castle, D.J., 2016. Comparing the experience of voices in borderline personality disorder with the experience of voices in a psychotic disorder: a systematic review. Aust. N. Z. J. Psychiatry 50 (7), 640–648. https://doi.org/10.1177/0004867416632595.
Moseley, P., Aleman, A., Allen, P., Bell, V., Bless, J., Bortolon, C., et al., 2020a. Correlates of hallucinatory experiences in the general population: an international multi-site replication study. Psychol. Sci. 32 (7), 1027–1037. https://doi.org/10.1177/0956797620985832.
Moseley, P., Alderson-Day, B., Common, S., Dodgson, G., Lee, R., Mitrenga, K.J., et al., 2020b. Continuities and Discontinuities in the Cognitive Mechanisms Associated with Clinical and Non-Clinical Auditory Verbal Hallucinations.
Mou, X., Bai, F., Xie, C., Shi, J., Yao, Z., Hao, G., et al., 2013. Voice recognition and altered connectivity in schizophrenic patients with auditory hallucinations. Prog. Neuro-Psychopharmacol. Biol. Psychiatry 44, 265–270. https://doi.org/10.1016/j.pnpbp.2013.03.006.
Mulert, C., Jäger, L., Propp, S., Karch, S., Störmann, S., Pogarell, O., et al., 2005. Sound level dependence of the primary auditory cortex: simultaneous measurement with 61-channel EEG and fMRI. NeuroImage 28 (1), 49–58. https://doi.org/10.1016/j.neuroimage.2005.05.041.
Nayani, T.H., David, A.S., 1996. The auditory hallucination: a phenomenological survey. Psychol. Med. 26 (1), 177–189. https://doi.org/10.1017/S003329170003381X.
Peters, E., Ward, T., Jackson, M., Woodruff, P., Morgan, C., McGuire, P., Garety, P.A., 2017. Clinical relevance of appraisals of persistent psychotic experiences in people with and without a need for care: an experimental study. Lancet Psychiatry 4 (12), 927–936. https://doi.org/10.1016/S2215-0366(17)30409-1.
Pinheiro, A.P., Rezaii, N., Rauber, A., Nestor, P.G., Spencer, K.M., Niznikiewicz, M., 2017. Emotional self–other voice processing in schizophrenia and its relationship with hallucinations: ERP evidence. Psychophysiology 54 (9), 1252–1265. https://doi.org/10.1111/psyp.12880.
Pinheiro, A.P., Schwartze, M., Kotz, S.A., 2018. Voice-selective prediction alterations in nonclinical voice hearers. Sci. Rep. 8 (1), 1–10. https://doi.org/10.1038/s41598-018-32614-9.
Pinheiro, A.P., Schwartze, M., Kotz, S.A., 2020. Cerebellar circuitry and auditory verbal hallucinations: an integrative synthesis and perspective. Neurosci. Biobehav. Rev. 118, 485–503. https://doi.org/10.1016/j.neubiorev.2020.08.004.

Plaze, M., Bartrés-Faz, D., Martinot, J., Januel, D., Bellivier, F., De Beaurepaire, R., et al., 2006. Left superior temporal gyrus activation during sentence perception negatively correlates with auditory hallucination severity in schizophrenia patients. Schizophr. Res. 87 (1–3), 109–115. https://doi.org/10.1016/j.schres.2006.05.005.
Poldrack, R.A., Fletcher, P.C., Henson, R.N., Worsley, K.J., Brett, M., Nichols, T.E., 2008. Guidelines for reporting an fMRI study. NeuroImage 40 (2), 409–414. https://doi.org/10.1016/j.neuroimage.2007.11.048.
Poldrack, R.A., Baker, C.I., Durnez, J., Gorgolewski, K.J., Matthews, P.M., Munafò, M.R., et al., 2017. Scanning the horizon: towards transparent and reproducible neuroimaging research. Nat. Rev. Neurosci. 18 (2), 115. https://doi.org/10.1038/nrn.2016.167.
Puce, A., Hämäläinen, M.S., 2017. A review of issues related to data acquisition and analysis in EEG/MEG studies. Brain Sci. 7 (6), 58. https://doi.org/10.3390/brainsci7060058.
Rapin, L.A., Dohen, M., Lœvenbruck, H., Whitman, J.C., Metzak, P.D., Woodward, T.S., 2012. Hyperintensity of functional networks involving voice-selective cortical regions during silent thought in schizophrenia. Psychiatry Res. Neuroimaging 202 (2), 110–117. https://doi.org/10.1016/j.pscychresns.2011.12.014.
Rolland, B., Jardri, R., Amad, A., Thomas, P., Cottencin, O., Bordet, R., 2014. Pharmacology of hallucinations: several mechanisms for one single symptom? Biomed Res. Int. 2014, 307106 https://doi.org/10.1155/2014/307106.
Romme, M., Escher, S., Dillon, J., Corstens, D., Morris, M., 2009. Living With Voices: 50 Stories of Recovery: PCCS Books.
Rossell, S.L., Boundy, C.L., 2005. Are auditory-verbal hallucinations associated with auditory affective processing deficits? Schizophr. Res. 78 (1), 95–106. https://doi.org/10.1016/j.schres.2005.06.002.
Rossell, S.L., Van Rheenen, T.E., Groot, C., Gogos, A., O'Regan, A., Joshua, N.R., 2013. Investigating affective prosody in psychosis: a study using the Comprehensive Affective Testing System. Psychiatry Res. 210 (3), 896–900. https://doi.org/10.1016/j.psychres.2013.07.037.
Sambataro, F., Blasi, G., Fazio, L., Caforio, G., Taurisano, P., Romano, R., et al., 2010. Treatment with olanzapine is associated with modulation of the default mode network in patients with Schizophrenia. Neuropsychopharmacology 35 (4), 904–912. https://doi.org/10.1038/npp.2009.192.
Sanjuan, J., Lull, J.J., Aguilar, E.J., Martí-Bonmatí, L., Moratal, D., Gonzalez, J.C., et al., 2007. Emotional words induce enhanced brain activity in schizophrenic patients with auditory hallucinations. Psychiatry Res. Neuroimaging 154 (1), 21–29. https://doi.org/10.1016/j.pscychresns.2006.04.011.
Scheinost, D., Tokoglu, F., Hampson, M., Hoffman, R., Constable, R.T., 2019. Data-driven analysis of functional connectivity reveals a potential auditory verbal hallucination network. Schizophr. Bull. 45 (2), 415–424. https://doi.org/10.1093/schbul/sby039.
Scherer, K.R., 2018. Acoustic patterning of emotion vocalizations. In: Belin, P., Frühholz, S. (Eds.), The Oxford Handbook of Voice Perception. Oxford University Press, Oxford, UK.
Shergill, S.S., Murray, R.M., McGuire, P.K., 1998. Auditory hallucinations: a review of psychological treatments. Schizophr. Res. 32 (3), 137–150. https://doi.org/10.1016/S0920-9964(98)00052-8.
Silverman, D.H.S., 2004. Brain 18F-FDG PET in the diagnosis of neurodegenerative dementias: comparison with perfusion SPECT and with clinical evaluations lacking nuclear imaging. J. Nucl. Med. 45 (4), 594–607.
Slade, P.D., Bentall, R.P., 1988. Sensory Deception: a Scientific Analysis of Hallucination. Johns Hopkins University Press.
Steinmann, S., Leicht, G., Ertl, M., Andreou, C., Polomac, N., Westerhausen, R., et al., 2014a. Conscious auditory perception related to long-range synchrony of gamma oscillations. NeuroImage 100, 435–443. https://doi.org/10.1016/j.neuroimage.2014.06.012.
Steinmann, S., Leicht, G., Mulert, C., 2014b. Interhemispheric auditory connectivity: structure and function related to auditory verbal hallucinations. Front. Hum. Neurosci. 8, 55. https://doi.org/10.3389/fnhum.2014.00055.
Steinmann, S., Leicht, G., Andreou, C., Polomac, N., Mulert, C., 2017. Auditory verbal hallucinations related to altered long-range synchrony of gamma-band oscillations. Sci. Rep. 7 (1), 1–10. https://doi.org/10.1038/s41598-017-09253-7.
Steinmann, S., Leicht, G., Mulert, C., 2019. The interhemispheric miscommunication theory of auditory verbal hallucinations in schizophrenia. Int. J. Psychophysiol. 145, 83–90. https://doi.org/10.1016/j.ijpsycho.2019.02.002.
Stephane, M., Burton, P., Meriwether, D., Yoon, G., 2018. Other" tags for "Self"-generated speech in patients with auditory verbal hallucinations, an fMRI study. Schizophr. Res. 202, 410–411. https://doi.org/10.1016/j.schres.2018.06.056.
Tagliazucchi, E., Von Wegner, F., Morzelewski, A., Brodbeck, V., Laufs, H., 2012. Dynamic BOLD functional connectivity in humans and its electrophysiological correlates. Front. Hum. Neurosci. 6 (339) https://doi.org/10.3389/fnhum.2012.00339.
Thiebes, S., Steinmann, S., Curic, S., Polomac, N., Andreou, C., Eichler, I.C., et al., 2018. Alterations in interhemispheric gamma-band connectivity are related to the emergence of auditory verbal hallucinations in healthy subjects during NMDA-receptor blockade. Neuropsychopharmacology 43 (7), 1608–1615.
Toh, W.L., Tan, E.J., Neill, E., Van Rheenen, T.E., Gurvich, C., Sumner, P.J., et al., 2020a. Identifying the cognitive underpinnings of voice-hearing by comparing never, past and current voice-hearers. Acta Psychiatr. Scand. 141 (6), 553–562. https://doi.org/10.1111/acps.13169.
Toh, W.L., Thomas, N., Hollander, Y., Rossell, S.L., 2020b. On the phenomenology of auditory verbal hallucinations in affective and non-affective psychosis. Psychiatry Res. 290, 113147 https://doi.org/10.1016/j.psychres.2020.113147.
Tovar, A., Fuentes-Claramonte, P., Soler-Vidal, J., Ramiro-Sousa, N., Rodriguez-Martinez, A., Sarri-Closa, C., et al., 2019. The linguistic signature of hallucinated voice talk in schizophrenia. Schizophr. Res. 206, 111–117. https://doi.org/10.1016/j.schres.2018.12.004.
Uhlhaas, P.J., Singer, W., 2010. Abnormal neural oscillations and synchrony in schizophrenia. Nat. Rev. Neurosci. 11 (2), 100–113. https://doi.org/10.1038/nrn2774.
Upthegrove, R., Broome, M.R., Caldwell, K., Ives, J., Oyebode, F., Wood, S.J., 2016. Understanding auditory verbal hallucinations: a systematic review of current evidence. Acta Psychiatr. Scand. 133 (5), 352–367. https://doi.org/10.1111/acps.12531.
van Lutterveld, R., Diederen, K.M.J., Otte, W.M., Sommer, I.E., 2014. Network analysis of auditory hallucinations in nonpsychotic individuals. Hum. Brain Mapp. 35 (4), 1436–1445. https://doi.org/10.1002/hbm.22264.
Van Os, J., Reininghaus, U., 2016. Psychosis as a transdiagnostic and extended phenotype in the general population. World Psychiatry 15 (2), 118–124. https://doi.org/10.1002/wps.20310.
Waters, F., Fernyhough, C., 2017. Hallucinations: a systematic review of points of similarity and difference across diagnostic classes. Schizophr. Bull. 43 (1), 32–43. https://doi.org/10.1093/schbul/sbw132.
Woodruff, P.W.R., Wright, I.C., Bullmore, E.T., Brammer, M., Howard, R.J., Williams, S.C.R., et al., 1997. Auditory hallucinations and the temporal cortical response to speech in schizophrenia: a functional magnetic resonance imaging study. Am. J. Psychiatry 154 (12), 1676–1682. https://doi.org/10.1176/ajp.154.12.1676.
Woodward, T.S., Menon, M., Whitman, J.C., 2007. Source monitoring biases and auditory hallucinations. Cogn. Neuropsychiatry 12 (6), 477–494. https://doi.org/10.1080/13546800701307198.
Zhang, Z., Hao, G.F., Shi, J.B., Mou, X.D., Yao, Z.J., Chen, N., 2008a. Investigation of the neural substrates of voice recognition in Chinese schizophrenic patients with auditory verbal hallucinations: an event-related functional MRI study. Acta Psychiatr. Scand. 118 (4), 272–280. https://doi.org/10.1111/j.1600-0447.2008.01243.x.
Zhang, Z., Shi, J., Yuan, Y., Hao, G., Yao, Z., Chen, N., 2008b. Relationship of auditory verbal hallucinations with cerebral asymmetry in patients with schizophrenia: an event-related fMRI study. J. Psychiatr. Res. 42 (6), 477–486. https://doi.org/10.1016/j.jpsychires.2007.04.003.