Distinct processing of ambiguous speech in people with non-clinical auditory verbal hallucinations

Ben Alderson-Day,[1,]* César F. Lima,[2,3,]* Samuel Evans,[2,4] Saloni Krishnan,[2,5] Pradheep Shanmugalingam,[2] Charles Fernyhough[1,†] and Sophie K. Scott[2,†]

*,†These authors contributed equally to this work.

Auditory verbal hallucinations (hearing voices) are typically associated with psychosis, but a minority of the general population also experience them frequently and without distress. Such 'non-clinical' experiences offer a rare and unique opportunity to study hallucinations apart from confounding clinical factors, thus allowing for the identification of symptom-specific mechanisms. Recent theories propose that hallucinations result from an imbalance of prior expectation and sensory information, but whether such an imbalance also influences auditory-perceptual processes remains unknown. We examine for the first time the cortical processing of ambiguous speech in people without psychosis who regularly hear voices. Twelve non-clinical voice-hearers and 17 matched controls completed a functional magnetic resonance imaging scan while passively listening to degraded speech ('sine-wave' speech), that was either potentially intelligible or unintelligible. Voice-hearers reported recognizing the presence of speech in the stimuli before controls, and before being explicitly informed of its intelligibility. Across both groups, intelligible sine-wave speech engaged a typical left-lateralized speech processing network. Notably, however, voice-hearers showed stronger intelligibility responses than controls in the dorsal anterior cingulate cortex and in the superior frontal gyrus. This suggests an enhanced involvement of attention and sensorimotor processes, selectively when speech was potentially intelligible. Altogether, these behavioural and neural findings indicate that people with hallucinatory experiences show distinct responses to meaningful auditory stimuli. A greater weighting towards prior knowledge and expectation might cause non-veridical auditory sensations in these individuals, but it might also spontaneously facilitate perceptual processing where such knowledge is required. This has implications for the understanding of hallucinations in clinical and non-clinical populations, and is consistent with current 'predictive processing' theories of psychosis.

1 Department of Psychology, Durham University, Science Laboratories, South Road, Durham, DH1 3LE, UK
2 Institute of Cognitive Neuroscience, University College London, 17–19 Queen Square, London, WC1N 3AR, UK
3 Faculty of Psychology and Education Sciences, University of Porto, Rua Alfredo Allen, 4200-135 Porto, Portugal
4 Department of Psychology, University of Westminster, 115 New Cavendish Street, London, W1W 6UW, UK
5 Department of Experimental Psychology, University of Oxford, S Parks Rd, Oxford OX1 3UD, UK

Correspondence to: Dr Ben Alderson-Day,
Department of Psychology, Durham University, Science Laboratories, South Road, Durham, DH1 3LE, UK
E-mail: benjamin.alderson-day@durham.ac.uk

Introduction

Auditory verbal hallucinations are typically studied in the context of schizophrenia. However, the presence of other clinical factors, such as additional symptoms or the use of medication, makes it challenging to investigate neurocognitive mechanisms that are hallucination-specific. One solution is to study auditory verbal hallucinations—or more commonly 'voice-hearing'—in the minority of the general population who have such experiences without need for care (Johns *et al.*, 2014). The existence of 'non-clinical' voice-hearing has been noted for many years and is strongly argued for by community groups (Romme and Escher, 1989; Corstens *et al.*, 2014). Estimates for voice-hearing in the general population vary from 5% to 15% (Beavan *et al.*, 2011), but rates for frequent and complex voices appear closer to 1–2% (Johns *et al.*, 1998; Kråkvik *et al.*, 2015). Such non-clinical voice-hearing (NCVH) is featurally similar to auditory verbal hallucinations described in psychosis, but usually more controllable and positive in content (Daalman *et al.*, 2011). Many non-clinical voice-hearers value their experiences and may seek to cultivate them over time (Baumeister *et al.*, 2017; Powers *et al.*, 2017).

Concerns about stigma make the recruitment of non-clinical voice-hearers extremely challenging: consequently, only a handful of studies have sought to examine the neurocognitive features of NCVH (Linden *et al.*, 2011; Kompus *et al.*, 2013). The most successful of these was conducted in Utrecht, Holland, which initially identified 103 people with frequent NCVH who did not qualify for a psychiatric diagnosis (Sommer *et al.*, 2010). To date, this remains the only project to have carried out neuroimaging studies in NCVH samples greater than 10 (Diederen *et al.*, 2012; de Weijer *et al.*, 2013; van Lutterveld *et al.*, 2014). These studies have shown that when hearing voices, people with NCVH and clinical auditory verbal hallucinations engage similar brain networks associated with speech and language processing, including the bilateral superior temporal gyrus (STG), inferior frontal gyrus (IFG) and anterior insula (Diederen *et al.*, 2012). The experience of NCVH likely also involves regions associated with the generation and monitoring of speech-motor imagery, as well as sensorimotor processes, such as the supplementary and pre-supplementary motor areas (SMA/pre-SMA; Linden *et al.*, 2011; Lima *et al.*, 2016). Atypical modulation of sensory cortex, by attention/monitoring and sensorimotor processes in the SMA/pre-SMA and adjacent anterior cingulate cortex (ACC), has been proposed as a potential mechanism underlying the experience of auditory verbal hallucinations (Allen *et al.*, 2007).

In behavioural studies, people with NCVH appear to be particularly susceptible to semantic expectation effects when instructed to monitor for speech in white noise (Daalman *et al.*, 2012), a result similar to effects seen in clinical voice-hearers and members of the general population who report milder, hallucination-like experiences (Fernyhough *et al.*, 2007; Vercammen *et al.*, 2008; Vercammen and Aleman, 2010; Varese *et al.*, 2012). Such effects have been interpreted as evidence of a bias in the perceptual processing of people with NCVH: a prior expectation for linguistic, meaningful percepts that would be sufficient to propagate internally-generated representations (e.g. speech imagery) down through speech and language networks, leading to non-veridical speech perception (Vercammen and Aleman, 2010; Daalman *et al.*, 2012).

However, if such 'priors for speech' are the mechanism underlying NCVH, their influence could be evident not just in speech monitoring tasks but also in speech processing more broadly, particularly when speech perception depends upon prior knowledge to disambiguate a degraded signal. An atypically strong prior for speech could actually facilitate processing, either spontaneously (allowing the hearer to identify potentially meaningful signals more easily) or when specifically directed by instructions (in turn enhancing the discrimination of speech from non-speech). This is consistent with recent evidence reported by Teufel *et al.* (2015) for visual processing in psychosis. People with an 'at risk' mental state (i.e. in early stages of psychosis) outperformed controls in their ability to identify objects in ambiguous, Mooney-style visual stimuli (Mooney, 1957), but only once they were given priming information about the objects. That is, people with hallucinations gained more from prior knowledge that could modulate their sensory predictions, leading to better skills in drawing meaning from noise. A similar effect in voice-hearers has never been demonstrated for the auditory domain, but can be tested using an ambiguous auditory stimulus: sine-wave speech (SWS).

SWS is a form of acoustically degraded speech, derived by synthesizing tones that track the amplitude and frequency of speech formants (Remez *et al.*, 1981). This can be used to produce potentially intelligible and unintelligible stimuli, based on whether the frequency and amplitude are drawn from the same or different original sentences (Rosen *et al.*, 2011). SWS is typically unintelligible on first exposure and may not be noticed as being speech-like (often sounding like 'aliens' or birdsong). Once the listener knows that it is potentially intelligible, though, relatively high levels of comprehension can be achieved (Remez *et al.*, 2011; Rosen *et al.*, 2011). Following training, SWS engages a left-lateralized 'speech mode' network including anterior and posterior temporal cortex (STG and middle temporal gyrus), IFG and insula (Vouloumanos *et al.*, 2001; Dehaene-Lambertz *et al.*, 2005; Benson *et al.*, 2006; Möttönen *et al.*, 2006; McGettigan *et al.*, 2012). Effects of prior knowledge and training on the processing of SWS and similar stimuli are reflected in the greater involvement of inferior frontal cortex (Davis and Johnsrude, 2003), pre-SMA, and dorsolateral prefrontal cortex (Eisner *et al.*, 2010; Rosen *et al.*, 2011), while posterior temporal cortex appears to track changes in sensory detail (Sohoglu *et al.*, 2012) and predictability (Gagnepain *et al.*, 2012).

Here we used SWS to study whether potential priors for speech in NCVH modulate the spontaneous processing of ambiguous sounds. NCVH participants and matched non-voice-hearing controls passively listened to intelligible and unintelligible SWS while being scanned in functional MRI, in a paradigm adapted from a study by Shanmugalingam *et al.* (2012). To disguise the presence of speech, participants were instructed to listen for a target cue (an equivalent noise-vocoded, unintelligible SWS stimulus, which sounded 'noisier' and 'rougher'), and were told that the other sounds (intelligible and unintelligible SWS) were 'distractor' stimuli (Fig. 1). After 20 min of scanning (Run 1), participants were asked if they had noticed any words or sentences in the distractor stimuli, and if so, when this occurred during the scan (visual markers were displayed during scanning to assist this, e.g. Block 1, 2 etc.). Participants were then explicitly told that there was actually speech in some of the stimuli (the 'reveal'), were trained to understand the SWS sentences within the scanner, and the scan was repeated, with the same set of stimuli and instructions (Run 2). After scanning, we tested the ability of participants to discriminate between intelligible SWS and unintelligible SWS (d'), their bias in classifying speech and non-speech (β), and accuracy (number of key words correct).

We anticipated that voice-hearers would show an enhanced ability to identify intelligible information in SWS when it was present, and our design allowed us to explore when and how this occurred. Behaviourally, if voice-hearers had a pre-existing prior for linguistic percepts, then this could be evident in an earlier recognition point for spontaneously identifying speech in the SWS stimuli. Alternatively, if voice-hearers were more likely to respond to the stimuli as speech-like only when their prior expectation for speech was explicitly modulated (following the reveal and training), this would result in no differences in recognition point, but potentially greater behavioural discrimination of speech and non-speech in the post-scanner task.

Neurally, potentially enhanced predictive representations of speech would be evident in a greater involvement of regions associated with prior knowledge effects on speech
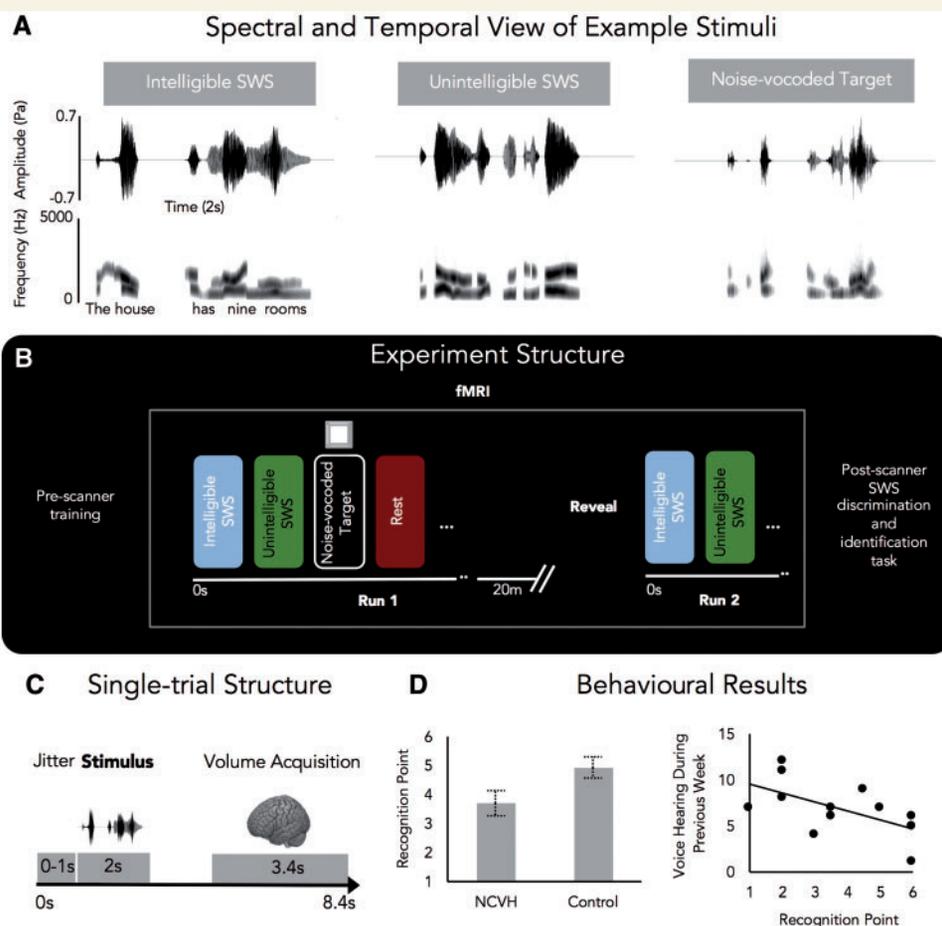
Figure 1 Experimental design and behavioural results. Participants were scanned in functional MRI while (**A**) listening to intelligible SWS, unintelligible SWS, or noise-vocoded, unintelligible target sounds; (**B**) listening and rest trials were presented in a pseudo-random order across two 20-min runs, divided by a 'reveal' period including training to understand SWS stimuli; (**C**) each trial lasted 8.4 s, including jitter, a 2 s stimulus and 3.4 s of volume acquisition; (**D**) NCVH participants recognized speech being present earlier than control participants during Run 1 (*left*), and this correlated with voice-hearing during the previous week (PSYRATS – Physical Characteristics subscale). PSYRATS = Psychotic Symptoms Rating Scale.

perception, including left inferior frontal cortex, pre-SMA and adjacent areas. If this reflected a spontaneous mechanism, then it would be seen before the reveal, and potentially also after; in other words, a general enhancement of the intelligibility response would be evident for NCVH participants. Alternatively, if it required explicit modulation, it would result in an enhancement of the intelligibility response only after the reveal. Both possibilities stand in contrast to the notion that the effect would be driven by differences in low-level auditory processes alone: a low-level effect (contrary to our expectations) would be evident in differential activation of sensory cortical regions (primary auditory cortex) across groups.

Materials and methods

Participants

The study included 12 NCVH participants and 17 non-voice-hearing control participants, matched for age, sex, handedness, education, and National Adult Reading Test scores (Nelson, 1982) (Table 1). All participants were aware that the study involved voice-hearers, but the project was described as focusing on 'how the brain processes unusual sounds', with study materials making no other reference to voices or speech.

Non-clinical voice-hearers were recruited in response to an online article for a national newspaper (Alderson-Day, 2014) and via social media, word of mouth, adverts with spiritual organizations, and previous participation in a related project (*n* = 4; the UNIQUE project; Peters *et al.*, 2016). Participants were included if they were over 18, had never received a psychiatric diagnosis in relation to voice-hearing, and endorsed any of three items derived from the revised Launay-Slade Hallucination Scale (LSHS; Bentall and Slade, 1985; Morrison *et al.*, 2000): 'In the past I have had the experience of hearing a person's voice that other people could not hear', 'I have heard a voice on at least one occasion in the past month',

or 'I have been troubled by hearing voices in my head'. Following Sommer *et al.* (2010), a phone screener was used to establish that (i) voices were distinct from thoughts and had a 'hearing quality'; (ii) voices were experienced at least once a month; (iii) voices were unrelated to drug or alcohol abuse; and (iv) no psychiatric diagnosis or treatment other than anxiety or depression in remission. Over an 18-month recruitment period, this identified 12 individuals who were then interviewed in more detail about their experiences (either at the participant's home or at a university location) and completed a functional MRI scanning session (see Supplementary material for interview details). Home visits were necessary due to the large geographical spread of participants across the UK.

Stimuli

The SWS stimuli were drawn from a stimulus set developed by Rosen *et al.* (2011) and used in McGettigan *et al.* (2012). Intelligible SWS and unintelligible SWS were identical to those previously used apart from being further noise-vocoded (Shannon *et al.*, 1995), a step we deliberately omitted to make them less noticeably speech-like. The only exception was the 'target' sounds, which were created by noise-vocoding a subset of 10 unintelligible SWS to change their timbre and make them distinctive from other stimuli. All SWS stimuli were derived from Bamford-Kowal-Bench sentences (e.g. 'The clown had a funny face'; Bench *et al.*, 1979) and recorded by an adult male speaker of standard Southern British English in an anechoic chamber. Frequency and amplitude from the first two formant tracks of each sentence were tracked and modelled with a sine wave tone using a semi-automatic procedure in MATLAB (The Mathworks, Natick, MA). Tracks were reviewed and hand-edited using custom software to ensure accurate tracking (Remez *et al.*, 2011; Rosen *et al.*, 2011). See Supplementary material for full details of the SWS preparation methods.

Pre-scan training

All training was conducted without mention of 'voices' or 'speech'. Participants were told that they would be listening

Table 1 Participant demographic and clinical characteristics

	NCVH		Control		P
Sex	8 F/4 M		12 F/5 M		0.822
Handedness	11 R/1 L		14 R/3 L		0.474
	Mean	**SD**	**Mean**	**SD**	
Age (years)	44.58	14.73	42.47	14.40	0.70
Education (years)	19.08	4.81	18.88	3.12	0.89
NART (max.50)	38.92	3.80	38.47	8.65	0.85
PSYRATS-AH Total	13.17	4.41	-	-	-
PSYRATS-AH 1-4 Interview	7.83	2.66	-	-	-
PSYRATS-AH 1-4 Scanning	6.92	2.97	-	-	-
PANSS-P	13.08	1.98	-	-	-
PANSS-N	8.00	0.95	-	-	-
P1 Delusions	2.33	0.78	-	-	-
P3 Hallucinations	4.00	0.60	-	-	-

F = female; M = male; R = right; L = left; NART = National Adult Reading Test; PSYRATS-AH = Psychotic Symptoms Rating Scale - Auditory Hallucinations; PANSS = Positive and Negative Syndrome Scale (P, Positive; N, Negative), P1 and P3 indicate individual PANSS items; higher ratings = greater severity. *P*-values correspond to chi-square tests for categorical data and two-tailed *t*-tests (*df* = 27) for continuous data.

to a range of sounds in the scanner, and instructed to listen out for a target sound that would sound 'different' or 'noisier' than the others. We did not provide information about the potential vocal/speech nature of the stimuli, and did not perform a pre-scan task to assess speech perception abilities, in order to ensure that participants remained naïve regarding our key manipulation, so that spontaneous responses to the stimuli could be examined in the scanner. Participants were played an example target sound three times over Sennheiser HD25 headphones, and then played three more examples of target sounds along with five non-vocoded unintelligible SWS stimuli, in a random order. Participants indicated with a button-press when they heard a target sound, and the stimulus set was repeated until participants could consistently discriminate targets from non-targets (no participant required the sequence to be repeated more than three times).

Functional MRI task

Participants listened to the SWS sounds across two identical runs of 20 min, broken up into six 'blocks' that were marked with a visually presented text stimulus (Block 1, Block 2, etc) (Fig. 1). Each run contained 45 intelligible SWS trials, 45 unintelligible SWS trials and 18 target sounds, presented quasi-randomly (one stimulus per trial). Target sounds and 19 silent trials were distributed such that they were presented regularly but unpredictably across the run, with no more than two trials from the same condition occurring sequentially. For each run they were instructed to listen closely for the target sounds and press a button each time one was heard.

After the first run, while still in the scanner, participants were asked the following questions: (1) Did you notice any words or sentences in the sounds you heard? (2) If so, do you know when you first noticed them? (3) Could you understand the words? and (4) Could you repeat any of the words?

For question 2, participants were asked to estimate when they first noticed that words were present, using the visual markers displayed periodically during the run. This was scored to the nearest block (1–6); for example, if someone reported hearing speech 'from the start of block 4 onwards', they would receive a 4. If participants specifically stated noticing halfway through a block, or were unsure but offered a range (e.g. 'some time around block 3 or block 4'), they were allocated a half score (e.g. 3.5, 4.5) in an attempt to be more precise. This score was then used as their individual 'recognition point' and treated as a continuous variable for subsequent analyses. Participants were then told that the first run included some potentially intelligible sentences in the non-target stimuli (the reveal), before being played six new intelligible SWS sentences. Participants were played each sentence once, asked to repeat any words they could back to the experimenter, showed a written presentation of the sentence, and then played the sentence two more times, along with the written presentation of the sentence. This combination of distorted auditory presentation and clear written feedback has previously been used to demonstrate effective intelligibility training effects on similar degraded stimuli (Davis et al., 2005). This process was repeated a maximum of twice (for all six sentences) to ensure that participants could decode the potentially intelligible SWS sentences in Run 2. The instructions for Run 2 were the same as run 1, i.e. participants were not instructed to pay attention to the now intelligible SWS sentences and instead to just listen for the target sounds.

Participants also completed two 5-min resting state scans before and after the passive listening run as part of a separate study.

Post-scan behavioural task

Following scanning, participants were played 50 SWS stimuli in a random order (25 intelligible SWS, 25 unintelligible SWS). For each stimulus, participants told an experimenter (i) if speech was present; and (ii) if so, what was being said. To check that participants could decode new sentences and not just recognize repeated sentences, 20% of the stimuli were new to the participants. Following prior studies, the main outcomes were 'keyword accuracy' (number of key words correctly identified in intelligible SWS), d' (sensitivity to speech versus non-speech), and β (bias in identifying speech as present or absent). The post-scanner task was self-paced and took ~15 min.

MRI acquisition

MRI scanning was completed on a 1.5 T Siemens Avanto using a 32-channel birdcage headcoil. Whole-brain echo-planar images were collected in two runs of 147 volumes each, using a sparse-sampling routine in which auditory stimuli were presented during the silent gap between brain acquisitions (Hall et al., 1999). The following parameters were used: repetition time = 8.4 s; acquisition time = 3.4 s, echo time = 0.5 s, flip angle = 90°, 40 axial slices, 3 mm³ in plane resolution. For localization, high resolution anatomical images were also acquired using a T_1-weighted magnetization prepared rapid acquisition gradient echo sequence (MP-RAGE; repetition time = 2.73 s, echo time = 3.57 ms, flip angle = 7°, 176 sagittal slices, voxel size = 1 mm³).

Auditory onsets occurred 5 s (±0–1-s jitter) before the beginning of the following volume acquisition. The stimuli were presented using Psychtoolbox (Brainard, 1997), running in MATLAB, via a Sony STR-DH510 digital AV control center (Sony) and MRI-compatible insert earphones (Sensimetrics Corporation). The sound volume was individually adjusted to a comfortable hearing level prior to scanning. All participants reported being able to hear the sounds without any difficulty.

MRI analysis

MRI analysis was conducted using Statistical Parametric Mapping software (SPM version 8; Wellcome Trust Centre for Neuroimaging, London, UK). The first two volumes of each run were discarded to allow longitudinal magnetization to ensure signal equilibrium. Functional images were realigned with the first volume per run and the anatomical T_1 image was then co-registered to the mean functional image. Functional images were then spatially normalized to MNI space using the parameters acquired from segmentation, resampled to 2 mm³ voxels, and smoothed using a Gaussian kernel of 8 mm³ full-width at half-maximum to ameliorate differences in intersubject localization. Responses for events of interest were modelled using a canonical haemodynamic response function. Intelligible SWS, unintelligible SWS, target sounds and visual stimuli (block titles) were modelled from their onsets with durations of 2 s, with silent trials acting as an implicit 'rest' baseline. Within each run, individual conditions were modelled as separate regressors in a generalized linear model (GLM), along with six movement parameters derived from realignment (three

translations, three rotations), that were included as regressors of no interest.

At the first-level (single-subject), T-contrast images were generated for the comparison of each of the conditions (intelligible SWS, unintelligible SWS, vigilance targets) against the implicit rest baseline. The following planned contrasts were also generated during first-level analyses:

(i) (intelligible SWS Run 1 + intelligible SWS Run 2) − (unintelligible SWS Run 1 + unintelligible SWS Run 2), corresponding to the general effect of intelligibility across runs. If NCVH participants spontaneously responded to intelligible stimuli in a distinct manner, group differences would be expected for this contrast.

(ii) (intelligible SWS Run 2 − unintelligible SWS Run 2) − (unintelligible SWS Run 1 − intelligible SWS Run 1), corresponding to a larger intelligibility response on Run 2 versus Run 1, once intelligible SWS were explicitly revealed as speech and participants were trained to understand it. If explicit modulation of expectations was required to trigger a distinct processing of intelligible stimuli in NCVH participants, group differences would be expected for this contrast.

(iii) intelligible SWS Run 1 − unintelligible SWS Run 1, corresponding to the intelligibility response prior to the reveal. Finding group differences for this contrast would further support the argument that NCVH spontaneously respond to intelligible stimuli in a distinct manner, and it would establish that the reveal and training are not required for group differences to emerge.

(iv) intelligible SWS Run 2 − unintelligible SWS Run 2, corresponding to the intelligibility response post-reveal. Group differences could also be seen for this contrast, but would not directly establish or refute differences in spontaneous processing as participants had already been told about the existence of speech in the intelligible SWS.

These images were taken up to second-level random effects analyses for group inferences. Where group differences were observed, analyses were repeated controlling for any behavioural differences between the groups (i.e. a difference in recognition point) by including them as covariates in the second-level analyses. We also carried out exploratory individual differences analyses in SPM, to examine associations between neural responses and behavioural performance. All statistical maps were thresholded at $P < 0.001$ peak-level uncorrected, cluster corrected with a family-wise error (FWE) at $P < 0.05$ across the whole-brain. All co-ordinates are reported in MNI space. Anatomical labels are based on the SPM Anatomy toolbox (Eickhoff *et al.*, 2005) and the Human Motor Area Template (HMAT; Mayka *et al.*, 2006), with images produced using SPM and MRIcroGL. Parameter estimates were extracted for plotting using the MarsBaR toolbox (Brett *et al.*, 2002) with regions of interest based on the full cluster extent of activated regions in the above analyses. Between-groups comparison of behavioural data was analysed using two-tailed *t*-tests at $P < 0.05$, unless otherwise specified.

Results

Behavioural analyses

During the training phase, some participants described the sounds as being 'a bit like a robot' or 'like the Clangers',

but no participants described either the target or unintelligible SWS sounds as being speech or voice-like. However, while being scanned, the majority of NCVH participants reported perceiving speech in the SWS stimuli before the mid-scan reveal, with one participant reporting hearing speech from the first 'three or four words' of Run 1. A significant difference was evident for the recognition point when participants reported first noticing words in the SWS: on average, the NCVH group heard them a block earlier than controls, as shown in Fig. 1D [mean = 3.71 and 4.94 for NCVH and control participants, respectively; $t(27) = -2.17$, $P = 0.039$]. [Due to non-normal data in the control group this comparison was also run using a permutation test in the *perm* package for R, producing similar results (mean difference = -1.23, $P = 0.041$), Monte Carlo method used with 2000 replications.] Overall, 9/12 NCVH participants (75%) reported realizing that there were words present compared to only 8/17 controls (47%). Of these, seven NCVH and five control participants additionally mentioned that they could understand the words, with five in each group being able to accurately recall some of them.

During scanning, all participants remained awake and responsive to the target stimuli, as indicated by the button-press data. However, button-press responses for four participants (one participant with NCVH, three control subjects) did not record correctly and one NCVH participant accidentally pressed a button for every trial. There were no group differences in total button presses, whether or not the latter participant was included (all $t < 1.4$, all $P > 0.19$). Participants with irregular button-press data were marked and checked for their influence on group comparisons of functional MRI data (see below). Only one NCVH participant reported experiencing a hallucination during scanning (a visual hallucination, occurring midway through Run 2); however, they did not report this affecting their ability to complete the task.

On the post-scan behavioural task (i.e. after all participants had been trained to understand the SWS sentences), no differences were observed between the groups, with similar performance for speech discrimination (*d'*), the ability to comprehend intelligible SWS (keyword accuracy), and bias to classify stimuli as speech (*β*; Supplementary Table 2).

Functional MRI

Responses to intelligible and unintelligible sine-wave speech over rest

Compared to rest, responses to intelligible (Fig. 2A) and unintelligible (Fig. 2B) SWS activated an extensive bilateral fronto-temporo-parietal network, including primary auditory cortex, IFG, SMA, inferior parietal lobule, and posterior STG. No supra-threshold group differences were evident for either the combination of intelligible and unintelligible SWS versus rest (i.e. the main effect of group during listening to sounds), nor any simple effects (i.e. the
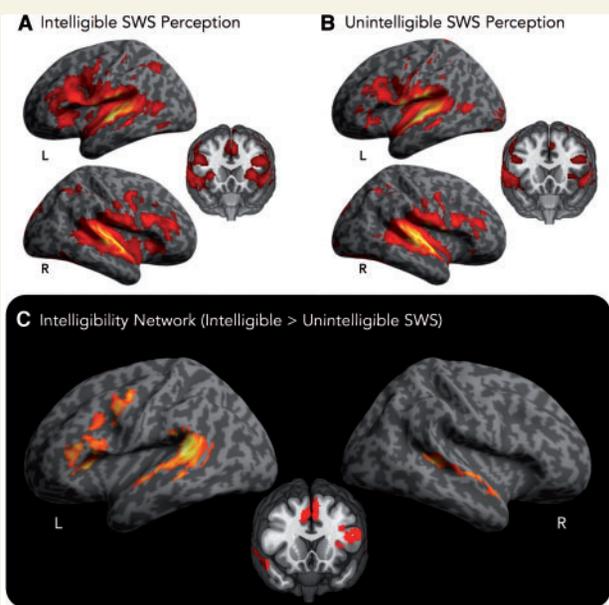
Figure 2 Responses to SWS and intelligibility effect across groups. Responses versus rest baseline to intelligible SWS (**A**), unintelligible SWS (**B**), and the difference between them, i.e. the intelligibility effect (**C**). Activation maps are presented at an uncorrected threshold of $P < 0.001$ peak level, FWE corrected ($P < 0.05$) at cluster level.

main effect of group during listening to intelligible-only SWS versus rest and unintelligible-only SWS versus rest).

Intelligibility effect

Across both runs and groups, several regions were more active for intelligible compared to unintelligible SWS, including the left and right STG, the left middle temporal gyrus, insula, precentral gyrus and IFG, as well as medial regions, namely the pre-SMA, ACC, and superior frontal gyrus (Table 2 and Fig. 2C). Between-groups comparisons of the intelligibility response [Intelligible > Unintelligible SWS, planned contrast (i)] indicated that NCVH participants showed greater activation than controls in a cluster with peaks in rostral ACC, extending to the pre-SMA, middle cingulate cortex, and superior frontal gyrus (Table 2 and Fig. 3C). That is, NCVH showed an enhanced discrimination between intelligible and unintelligible SWS within these regions. Plotting the response of this cluster indicated that the effect was mostly driven by increased responses to intelligible SWS in NCVH (Fig. 3C, right). To further test this observation, we directly compared the groups' beta values for this cluster within each SWS condition: voice-hearers showed significantly greater responses than controls for intelligible SWS [$t(27) = 2.98$, $P = 0.006$], but the groups were similar for unintelligible SWS [$t(27) = -1.05$, $P = 0.301$]. The reverse contrast (Controls > NCVH) yielded no significant clusters.

As some participants reported hearing speech before the reveal, it could be that group differences evident in the intelligibility response simply reflected NCVH participants

having more opportunity to listen to intelligible SWS in 'speech mode'. To examine this, we reran the group comparison of Intelligible > Unintelligible SWS with the timing of participants' noticing of speech—their recognition point—included as a covariate. The group difference in ACC remained significant (MNI coordinates for peak voxel: -2, 32, 26, $k = 467$, $t = 5.27$, $z = 4.31$, $P_{FWE} < 0.001$), indicating that greater recruitment of this region by NCVH participants was unlikely to simply reflect a confound resulting from an earlier switch to speech mode. We also confirmed that the pattern of findings remained unchanged when excluding the participant who pressed a button on every trial and those without a full record of button presses; as such, all participants were retained for the remainder of analyses.

The effect of the reveal: interaction between run and intelligibility

With the two groups combined, there was a significant interaction for the intelligibility response from Run 1 to Run 2 in left pSTG (MNI coordinates for peak voxel: -50, -48, 10, $k = 790$, $t = 5.65$, $z = 4.55$, $P_{FWE} < 0.001$; see Fig. 3D). This change was specific to intelligible stimuli, i.e. no effect was evident for the change in responses to unintelligible stimuli (Fig. 3D, right). This pattern was confirmed in a follow-up analysis, after extracting beta values for this cluster: responses to intelligible SWS were stronger in Run 2 than in Run 1 [$t(28) = -4.08$, $P < 0.001$], but responses to unintelligible SWS were similar across runs [$t(28) = 0.12$, $P = 0.909$].

There were no supra-threshold group differences for an interaction effect from Run 1 to Run 2 [i.e. planned contrast (ii)]. That is, NCVH participants did not show a specific benefit in intelligibility once trained to listen for speech, indicating that the effect of the reveal and subsequent training had a broadly similar influence on intelligibility responses across groups. Even with a more liberal threshold ($P < 0.001$ peak level, uncorrected), no clusters over 50 voxels were observed within grey matter.

In the separate analyses for Runs 1 and 2 [planned contrasts (iii) and (iv)] a clear intelligibility network was observed for Run 2 but not Run 1 for both groups (Table 3), consistent with the non-significant interaction observed. Contrary to what would be expected if group differences were dependent on the explicit modulation of expectation, NCVH already showed a stronger intelligibility response than controls in Run 1, in the same ACC region as in the overall analysis (MNI coordinates for peak voxel: 2, 36, 28, $k = 241$, $t = 5.14$, $P_{FWE} = 0.008$) and in left middle frontal gyrus (MNI coordinates for peak voxel: -36, 54, 0, $k = 190$, $t = 4.91$, $P_{FWE} = 0.024$). Group differences in ACC for intelligibility were also evident in Run 2, albeit at subthreshold levels (MNI coordinates for peak voxel: -4, 38, 20, $k = 51$, $t = 4.12$, $P < 0.001$ uncorrected), which was consistent with the general enhancement of an intelligibility effect across the whole scanning session.

Table 2 Responses to Intelligible > Unintelligible SWS, across both runs and groups combined, and group differences in intelligibility responses

Location	x	y	z	Voxels, n	t	z	P_{FWE}
Runs and groups combined							
L superior temporal gyrus	−62	−42	16	1763	8.24	5.78	<0.001
L superior temporal gyrus	−52	−44	20		7.99	5.67	
L middle temporal gyrus	−58	−34	4		6.39	4.95	
L middle temporal gyrus	−58	−16	−2		6.15	4.82	
L middle temporal gyrus	−60	−26	2		5.07	4.21	
L middle temporal gyrus	−56	−62	12		3.58	3.21	
L middle temporal gyrus	−48	−26	0		3.54	3.18	
L pre-supplementary motor area	−4	16	58	755	6.19	4.84	<0.001
R pre-supplementary motor area	6	18	46		5.58	4.51	
L pre-supplementary motor area	−4	20	44		4.99	4.16	
R anterior cingulate cortex	8	18	38		4.85	4.08	
L superior frontal gyrus	−6	28	44		4.39	3.78	
L precentral gyrus	−48	−2	48	1719	6.07	4.78	<0.001
L inferior frontal gyrus	−48	30	14		5.88	4.68	
L insula lobe	−32	24	8		5.81	4.64	
L inferior frontal gyrus	−46	16	16		5.34	4.38	
L precentral gyrus	−44	2	38		5.33	4.37	
L inferior frontal gyrus	−56	14	22		4.69	3.98	
L inferior frontal gyrus	−40	32	2		4.68	3.97	
L precentral gyrus	−42	6	30		4.37	3.77	
L inferior frontal gyrus	−36	14	30		4.26	3.70	
R superior temporal gyrus	62	−6	−4	468	5.71	4.58	<0.001
R superior temporal gyrus	60	−16	−2		4.77	4.03	
R superior temporal gyrus	54	14	−14		4.59	3.91	
R temporal pole	58	6	−12		4.57	3.90	
R superior temporal gyrus	52	−4	−14		3.72	3.31	
R superior temporal gyrus	48	−36	6	238	5.49	4.46	0.001
Group differences							
L anterior cingulate cortex	−4	34	26	539	4.81	4.05	<0.001
L middle cingulate cortex	−6	16	38		4.61	3.93	
R pre-supplementary motor area	6	14	42		4.61	3.92	
R middle cingulate cortex	8	22	36		4.40	3.78	
L anterior cingulate cortex	0	26	26		4.36	3.76	
L superior frontal gyrus	−12	28	32		4.03	3.53	
L superior frontal gyrus	−6	30	38		3.84	3.40	

These results are presented at an uncorrected threshold of $P < 0.001$ peak level, FWE corrected ($P < 0.05$) at cluster level. L = Left; R = Right. We report a maximum of 15 grey matter local maxima (that are >8 mm apart) per cluster.

Comparing responses in primary auditory cortex

The lack of supra-threshold group differences in responses to intelligible or unintelligible stimuli over rest indicated that basic auditory processes were broadly similar in controls and NCVH. To explore this further, we extracted average responses to intelligible and unintelligible sounds in the bilateral primary auditory cortices (defined as TE 1.0, 1.1, and 1.2 based on the SPM Anatomy Toolbox) and conducted Bayesian inference testing on effects of group, intelligibility, and run. A Bayesian mixed ANOVA was conducted using JASP (JASP Team, 2017), with the default priors (Rouder *et al.*, 2017). When a model containing the group effect was compared to one without it (i.e. the null hypothesis), the Bayes Factor was 0.54, or 1:1.86 in favour of the null (in other words, the data

were almost twice as likely to occur under the null hypothesis). Evidence for any group-related interaction effects was even weaker: Bayes Factor values of 0.26, 0.26 and 0.57 were observed for models containing group × run, group × intelligibility, and group × run × intelligibility, respectively (i.e. 1:3.85, 1:3.85 and 1:1.75 in favour of the null). Bayes Factor values for each model were calculated by comparing to the next most complex models lacking those terms (i.e. the three-way interaction model was compared with a model containing all two-way interactions; Rouder *et al.*, 2017). These values only reflect anecdotal to substantial evidence in favour of the null hypothesis (Jarosz and Wiley, 2014), but they nevertheless offer no evidence at all in favour of potential group differences in primary auditory cortex signal.
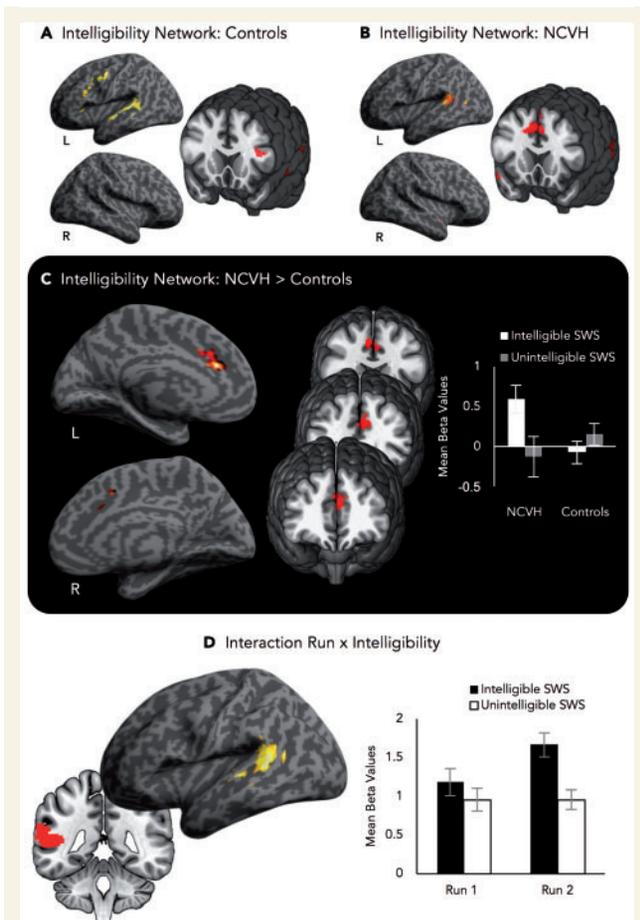
Figure 3 Group differences in intelligibility responses and effect of the reveal. Intelligibility responses in control participants (**A**), in voice-hearers (**B**), between-group differences in the intelligibility effect (**C**), and the change in the intelligibility effect following training with intelligible SWS, both groups combined (**D**). Beta values shown in (**C**) are extracted from a cluster with peak in the anterior cingulate cortex (MNI coordinates: −4, 34, 26) identified in whole-brain analysis. Beta values shown in (**D**) are extracted for a region of left STG (MNI coordinate: −50, −48, 10) identified in the Run × Intelligibility whole-brain interaction. Activation maps are presented at an uncorrected threshold of $P < 0.001$ peak level, FWE corrected ($P < 0.05$) at cluster level.

Individual differences in intelligibility responses

To explore how early responders may have been identifying speech in the SWS, we ran a whole-brain individual differences analysis, including recognition point as a regressor in the Intelligible > Unintelligible SWS contrast. The intelligibility response across Runs 1 and 2 in left IFG was negatively related to the recognition point (indicating that those who noticed speech earlier showed greater activation in these regions; Fig. 4A and Table 4). For Run 1 only (i.e. before all participants were in 'speech mode'), the recognition point was negatively related to responses in the middle cingulate cortex extending to parietal areas (Fig. 4B) and positively related to activation in medial prefrontal cortex (Fig. 4C). We also ran the same analysis for an index of

voice-hearing in the NCVH participants (PSYRATS Physical Characteristics from the past week; Haddock *et al.*, 1999; see Supplementary material); this indicated no significant whole-brain correlations. However, a behavioural correlation was observed between voice-hearing in the past week and recognition point (r = −0.582, *n* = 12, P = 0.047), such that a greater tendency to hear voices was associated with noticing speech earlier in Run 1 (Fig. 1D). This correlation directly links auditory-perceptual processes, as evaluated in the current study, with the magnitude of recent auditory verbal hallucinations.

Discussion

Despite decades of work on hallucinations, little is known about how they relate to everyday perceptual mechanisms. Our research aimed to address this by studying the interaction of expectation and perception in non-clinical voice-hearers. Knowledge and expectations help us to interpret ambiguous signals in a range of contexts; in some cases, this might lead to non-veridical sensations, but in other situations—such as hearing sine-wave speech—such expectations might contribute to divining meaningful signal from apparent noise (Davis and Johnsrude, 2007).

Behavioural evidence of NCVH hearing semantically congruent (but absent) speech in white noise (Daalman *et al.*, 2012) and signal detection biases in people prone to hallucinations (Brookwell *et al.*, 2013) has been used to argue for the existence of attentional factors—such as expectation and prior knowledge—having a greater influence on perception in people who hear voices. Our design, by initially disguising the presence of speech from participants, allowed us to examine whether such an influence can act spontaneously in NCVH, or requires the specific modulation of expectation (in essence, a suggestibility effect). The subjective behavioural responses of voice-hearers here—reporting the detection of speech content in the acoustics of SWS earlier than controls—suggest a spontaneous tendency in this group to extract meaningful linguistic information from ambiguous signals. Importantly, this finding is complemented by distinct responses seen in brain activity, as indicated by a stronger neural discrimination between intelligible and unintelligible SWS in NCVH participants. This effect could be seen even before the reveal and training, so was therefore not dependent on the modulation of expectation. Indeed, the comparable levels of discrimination and accuracy in the post-scanner task, and the absence of group differences in how the reveal and training affected brain responses, suggest that the explicit modulation of expectation does not play a major role in how NCVH process ambiguous speech.

This appears to contrast with the evidence reported by Teufel *et al.* (2015) that people with hallucinations benefit more from the modulation of prior knowledge, although both findings are potentially consistent with attention and expectation playing a role in unusual perceptions. Under

Table 3 Intelligibility responses (Intelligible > Unintelligible SWS) separately per run

Contrast	Location	x	y	z	Voxels, n	T	z	P_FWE
Run 1	-	-	-	-	-	-	-	-
Run 2	L inferior parietal cortex	−68	−42	20	2125	8.96	6.05	<0.001
	L superior temporal gyrus	−54	−44	14		8.22	5.77	
	L middle temporal gyrus	−60	−36	6		7.90	5.64	
	L middle temporal gyrus	−60	−26	2		6.55	5.02	
	L middle temporal gyrus	−58	−16	−2		5.47	4.45	
	L middle temporal gyrus	−44	−56	16		5.00	4.17	
	L middle temporal gyrus	−56	−62	14		4.60	3.92	
	L temporal pole	−58	8	−12		4.22	3.66	
	L superior temporal gyrus	−60	−2	−6		4.17	3.63	
	L middle temporal gyrus	−40	−64	18		4.13	3.60	
	L superior frontal gyrus	−6	28	44	965	6.49	5.00	<0.001
	Pre-supplementary motor area	0	20	46		6.25	4.87	
	L pre-supplementary motor area	−10	14	56		5.38	4.40	
	L pre-supplementary motor area	−2	10	60		5.06	4.21	
	R superior frontal gyrus	10	34	44		3.61	3.23	
	L insula lobe	−32	24	8	2220	6.37	4.93	<0.001
	L inferior frontal gyrus	−46	20	18		5.88	4.68	
	L inferior frontal gyrus	−42	24	0		5.80	4.63	
	L inferior frontal gyrus	−52	12	24		5.68	4.57	
	L precentral gyrus	−50	−4	46		5.08	4.22	
	L inferior frontal gyrus	−32	30	−10		5.08	4.22	
	L inferior frontal gyrus	−50	30	8		4.73	4.00	
	R superior temporal gyrus	64	−2	−6	868	6.18	4.84	<0.001
	R middle temporal gyrus	56	−36	6		6.04	4.76	
	R temporal pole	60	6	−12		5.42	4.42	
	R temporal pole	52	4	−14		5.35	4.38	
	R temporal pole	56	10	−18		5.19	4.28	
	R superior temporal gyrus	60	−16	−2		4.87	4.09	
	R superior temporal gyrus	44	−38	14		4.83	4.06	
	R inferior frontal gyrus	32	20	−2	216	5.52	4.48	0.019
	R inferior frontal gyrus	36	32	−6		4.22	3.74	

These results are presented at an uncorrected threshold of *P* < 0.001 peak level, FWE corrected (*P* < 0.05) at cluster level. L = Left; R = Right. We report a maximum of 15 grey matter local maxima (that are >8 mm apart) per cluster.

recent 'predictive processing' approaches (Clark, 2013), perception is understood as the balanced product of expectation-driven predictions (priors) about the external environment, and prediction error signals prompted by new sensory information (Rao and Ballard, 1999; Hohwy, 2014). Most predictive processing models—of hallucination specifically and psychosis more generally—posit a shift towards prior expectations, perhaps as a response to inherently unreliable prediction errors, or a top-down failure to modulate their precision (Grossberg, 2000; Friston, 2005; Fletcher and Frith, 2009; Adams *et al.*, 2013; Corlett *et al.*, 2016; Powers *et al.*, 2016). This is not always the case, however: the circular inference model (Jardri and Denève, 2013*a*), for example, proposes that hallucinations and delusions can result from an over-counting of sensory evidence instead, leading to a confusion of priors and prediction errors (see also Jardri and Denève, 2013*b*; Leptourgos *et al.*, 2015; Jardri *et al.*, 2016). Had our data only indicated a modulatory effect of the reveal on participants' responses, then it would have directly

supported an enhanced influence of new prior knowledge in the perceptual processing of NCVH (as in Teufel *et al.*, 2015). Instead, the spontaneous orientation towards speech that we observed could either be an indirect indicator of a pre-existing prior for speech, or be explained by differences in how the sensory signal is weighted. We did not observe significant group differences in primary sensory regions, either in whole-brain analysis or in follow-up Bayesian analysis. However, potential subtle differences in sensory weighting cannot be definitely ruled out using the present design. Further investigation of the intelligibility response using a paradigm that measures prior probability, sensory signal and participant response on a trial-by-trial basis would be required to examine this (for a recent example from decision-making, see Jardri *et al.*, 2017).
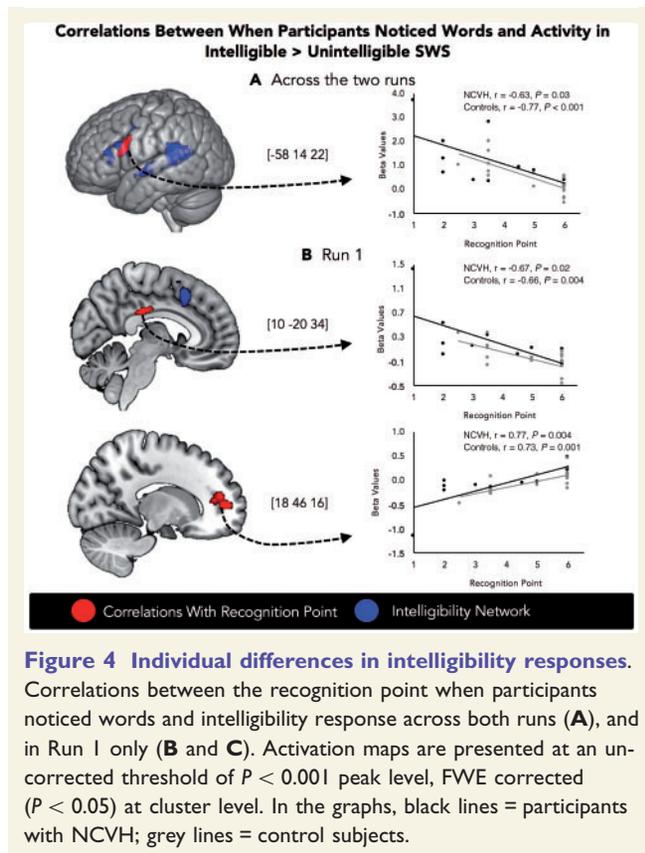
Given the subjective nature of our in-scanner 'recognition point' measure, the finding that group differences in the neural responses to SWS were specific to potentially intelligible signals is key. It suggests that NCVH were not simply biased to report perceiving speech in any signal,

and constrains the discussion of the potential mechanisms driving speech perception in voice-hearers. The lack of differences for any of the separate conditions versus rest, or any differences specific to primary auditory cortical regions, suggests that early auditory processes alone were unlikely to be driving group differences in intelligibility. However, speech areas that are usually associated with effects of prior knowledge and expectation—such as left inferior frontal cortex (Obleser and Kotz, 2010)—also showed no group differences. Instead, differences were seen in a region of

Figure 4 Individual differences in intelligibility responses. Correlations between the recognition point when participants noticed words and intelligibility response across both runs (**A**), and in Run 1 only (**B** and **C**). Activation maps are presented at an uncorrected threshold of $P < 0.001$ peak level, FWE corrected ($P < 0.05$) at cluster level. In the graphs, black lines = participants with NCVH; grey lines = control subjects.

rostral ACC, extending dorsally and caudally to reach the anterior pre-SMA and superior frontal gyrus.

Although part of the evolutionarily older midline vocalization network (Schulz et al., 2005), the ACC is not a classical speech processing area. Nevertheless, ACC responses have been observed for listening to distorted speech (Davis and Johnsrude, 2003), and ACC activation correlates with the accurate categorization of phonemes under adverse listening conditions (Du et al., 2014). In hallucinations research, the ACC has been associated with the monitoring and generation of internal and external speech (Simons et al., 2010), and linked to the occurrence of auditory verbal hallucinations, via atypical modulation of sensory regions (for a review see Allen et al., 2007). ACC activation has been observed during epochs of spontaneous activity in voice-selective areas of auditory cortex in healthy individuals (Hunter et al., 2006), 'self-induced' auditory hallucinations in hypnosis-prone people (Szechtman et al., 1998), and auditory attention in people with sleep-related hallucinations (Lewis-Hanna et al., 2011). ACC involvement was also observed in a number of early symptom-capture studies of people hearing voices while being scanned (Shergill et al., 2000), although later meta-analyses have failed to consistently identify this region during the hallucinatory state (Jardri et al., 2011; Kühn and Gallinat, 2012; Zmigrod et al., 2016).

The ACC is associated with a range of processes including attention, error monitoring, affect, and cognitive control (Devinsky et al., 1995). The dorsal, 'cognitive' ACC has been proposed to monitor task responses and attention, modulating selection bias and rule application in lateral prefrontal cortex (PFC) and inferior frontal cortex respectively (Langner and Eickhoff, 2013). Rostral areas of dorsal ACC appear sensitive to conflicts in response driven by irrelevant stimuli, while more caudal areas manage the allocation of attention (Orr and Weissman, 2009). The extension of this cluster into parts of pre-SMA is also notable given this area's prior implication in symptom-capture studies of auditory verbal hallucinations (Linden et al., 2011;

Table 4 Relationship between intelligibility responses (Intelligible > Unintelligible SWS) and the point at which participants reported recognizing that speech was present

Run	Location	x	y	z	Voxels, n	t	z	P_FWE
Runs 1 and 2	L inferior frontal gyrus	−58	14	22	189	4.81	4.05	0.038
	L precentral gyrus	−54	10	30		4.70	3.98	
	L inferior frontal gyrus	−46	30	12		4.46	3.83	
	L inferior frontal gyrus	−52	22	10		3.48	3.13	
Run 1	R middle cingulate cortex	10	−20	34	168	5.00	4.17	0.045
	L middle cingulate cortex	−4	−18	32		4.91	4.11	
	R superior parietal lobule	18	−34	34		3.82	3.38	
	L superior parietal lobule	−16	−30	36		3.74	3.32	
	R superior frontal gyrus	18	46	16	167	5.61	4.53	0.046
	R superior frontal gyrus	14	52	8		4.81	4.05	
	R anterior cingulate cortex	16	40	10		4.24	3.68	

These results are presented at an uncorrected threshold of $P < 0.001$ peak level, FWE corrected ($P < 0.05$) at cluster level. L = Left; R = Right. We report a maximum of 15 grey matter local maxima (that are >8 mm apart) per cluster.

Raij and Riekki, 2012), monitoring of inner speech (McGuire *et al.*, 1996), and the generation of sensorimotor predictions that guide and optimize perceptual processes (Lima *et al.*, 2016). The presence of dorsal ACC and pre-SMA together in the voice-hearer response may imply a greater attentional capture and sensorimotor processing of speech-like stimuli.

The individual difference results also provide clues as to how participants in both groups were able to identify speech in the SWS. Relationships between the recognition point when speech was noticed and activity in left IFG, medial PFC, and middle cingulate cortex (MCC) imply the involvement of both speech-motor processes and amodal, 'default mode' regions (Raichle *et al.*, 2001). The negative correlation with left IFG activation is consistent with the deployment of this region for parsing speech in adverse listening conditions, and may reflect the accessing of word meanings and segments to support perception via prior knowledge (Davis and Johnsrude, 2003; Obleser and Kotz, 2010; Sohoglu *et al.*, 2012; Du *et al.*, 2014). For instance, Eisner *et al.* (2010) found that the recruitment of the left IFG predicts individual differences in the listeners' ability to decode vocoded and spectrally shifted speech. Activity in the medial PFC, in contrast, is often linked with the default mode network (DMN) and would be consistent with participants taking longer to notice potentially intelligible SWS due to a lack of external engagement (Buckner *et al.*, 2008). The MCC cluster observed here is at the rostral border of the posterior cingulate cortex (PCC) and is sometimes classified as part of the dorsal subdivision of Brodmann area 23 (Cauda *et al.*, 2010), which is distinguished from ventral PCC regions posterior to the splenium (Vogt, 2016). Although the PCC and surrounding posterior midline structures are also associated with DMN-like task-negative activity, its dorsal subcomponents have been linked to networks responsible for cognitive control and external attention (Cauda *et al.*, 2010; Leech *et al.*, 2011; Leech and Sharp, 2014).

Some limitations of the present study must be acknowledged. First, for practical reasons—and because of the goals of the experiment—the behavioural assessment of participants' ability to discriminate and understand SWS had to be conducted outside the scanner and followed a long period of training and exposure to the stimuli. As such, it is possible that any post-scan group differences were masked or trained out as a result of the procedure, given that decoding of other kinds of degraded auditory stimuli—such as noise-vocoded speech—can improve over time and with training (Davis *et al.*, 2005). However, neither group performed at ceiling on the post-scan task: keyword accuracy after scanning was reasonably low in both groups compared to prior studies using distorted speech (McGettigan *et al.*, 2012), despite the fact that speech/non-speech discrimination was good. In future studies it will be important to assess NCVH participants' abilities

to decode SWS under a variety of listening conditions to measure decoding skill and adaptation more directly.

Second, we are reliant on the accuracy of participants' self-reports to gauge when participants noticed speech during Run 1, and cannot know for sure what participants were responding to when 'hearing' speech. Relying on self-report data is not uncommon in hallucinations research and retrospective reporting of events in the scanner has been used successfully to identify periods of voice-hearing (Jardri *et al.*, 2013). Nevertheless, it is possible that NCVH participants were just more likely to class any unusual stimuli as speech, rather than intelligible stimuli specifically. Two pieces of evidence militate against such an interpretation, though: first, the lack of any general group differences in the neural response to stimuli versus baseline (i.e. across both intelligible and unintelligible SWS), and second, the lack of any evident speech bias on the post-scan behavioural task. Notably, our brain data provide evidence in favour of a selective effect for the discrimination of intelligible stimuli: an effect that is hard to account by positing a non-specific response bias. Future studies could further address the selectivity of the behavioural effect by testing whether differences in recognition point also exist for a run without potentially intelligible SWS (this would be evidence for a non-specific bias), or by assessing degraded speech perception skills more comprehensively prior to training (e.g. Boebinger *et al.*, 2015). Including such conditions in the current study would have compromised our ability to test naïve participants' spontaneous responses to ambiguous stimuli.

Finally, we were restricted to a smaller sample of participants in the present study than is generally recommended for clinical functional MRI research (Carter *et al.*, 2008) and for group comparisons in general functional MRI studies (Poldrack *et al.*, 2017). Recruitment for neuroimaging studies with NCVH groups is extremely challenging: the present sample size is larger than other recent studies (Linden *et al.*, 2011; Kompus *et al.*, 2013), with the exception of the Utrecht cohort (Diederen *et al.*, 2012). Prior NCVH imaging studies have largely confined task-based functional MRI investigations to symptom capture (Linden *et al.*, 2011; Diederen *et al.*, 2012) or basic cognitive paradigms, such as dichotic listening (Kompus *et al.*, 2013) or verbal fluency (Diederen *et al.*, 2010), often with recourse to region of interest analysis and other methods of constraining analysis (and statistical corrections) to selected brain regions. To our knowledge, this is the first NCVH study to have successfully combined a complex behavioural paradigm with imaging data to examine a potential mechanism underlying hallucination, and while maintaining conservative whole-brain corrections. Nevertheless, small sample sizes in neuroimaging research with clinical and non-clinical voice-hearers is an enduring problem. As we have advocated elsewhere (Alderson-Day *et al.*, 2016) the combination of functional MRI data from multiple laboratories provides one means of addressing this issue. The International Consortium of Hallucinations Research

(ICHR) is currently supporting ongoing mega-analytic projects involving the combination of task-based, resting-state and structural MRI data from people with auditory verbal hallucinations (Thomas *et al.*, 2016).

Notwithstanding the small sample size of the present study, it is also important to note that the general response to intelligibility—and general effects of training with SWS—involved regions consistent with previous research on distorted speech. The primarily left-lateralized network seen across both groups is consistent with intelligibility effects using very similar stimuli (McGettigan *et al.*, 2012), as is the involvement of the SMA (Rosen *et al.*, 2011). The involvement of left posterior STG seen specifically following training also replicates prior findings using SWS (Möttönen *et al.*, 2006). Thus, in general, these two groups of participants showed plausible responses to the challenge of interpreting SWS.

In conclusion, the present study represents a first step in the understanding of atypical auditory-perceptual processes in people who regularly hear voices but do not require mental health support. Such individuals do not appear to be differentially affected by explicit modulations of expectation—instead, people in this group report being able to spontaneously extract speech from degraded auditory signals (and report doing so earlier than matched controls). This finding is broadly consistent with predictive processing models of hallucination and perception. The functional MRI results indicate that this capacity appears to rely less on enhanced speech-specific feedback to auditory regions, and more on the engagement of sensorimotor and domain-general attentional resources, selectively for potentially intelligible speech stimuli. This suggests that the fundamental mechanisms underlying hallucination involve—and may develop from—ordinary perceptual processes, illustrating the continuity of mundane and unusual experience. It has implications not only for 'continuum' views of experiences usually associated with psychosis (Johns and van Os, 2001), but also for the normalization, interpretation, and public understanding of a seriously misunderstood phenomenon.

Supplementary material

Supplementary material is available at *Brain* online.

References

Adams RA, Stephan KE, Brown HR, Frith CD, Friston KJ. The computational anatomy of psychosis. Front Psychiatry 2013; 4:47.

Alderson-Day B. Do you hear voices? You are not alone. In: The Guardian. 2014. Available from: https://www.theguardian.com/science/blog/2014/aug/13/hear-voices-auditory-verbal-hallucinations

Alderson-Day B, Diederen K, Fernyhough C, Ford JM, Horga G, Margulies DS, et al. Auditory hallucinations and the brain's resting-state networks: findings and methodological observations. Schizophr Bull 2016; 42:1110–23.

Allen P, Aleman A, Mcguire PK. Inner speech models of auditory verbal hallucinations: evidence from behavioural and neuroimaging studies. Int Rev Psychiatry 2007; 19:407–15.

Baumeister D, Sedgwick O, Howes O, Peters E. Auditory verbal hallucinations and continuum models of psychosis: a systematic review of the healthy voice-hearer literature. Clin Psychol Rev 2017; 51:125–41.

Beavan V, Read J, Cartwright C. The prevalence of voice-hearers in the general population: a literature review. J Ment Health 2011; 20:281–92.

Bench J, Kowal Å, Bamford J. The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. Br J Audiol 1979; 13:108–12.

Benson RR, Richardson M, Whalen DH, Lai S. Phonetic processing areas revealed by sinewave speech and acoustically similar non-speech. Neuroimage 2006; 31:342–53.

Bentall RP, Slade PD. Reliability of a scale measuring disposition towards hallucination: a brief report. Personal Individ Differ 1985; 6:527–9.

Boebinger D, Evans S, Scott SK, Rosen S, Lima CF, Manly T. Musicians and non-musicians are equally adept at perceiving masked speech. J Acoust Soc Am 2015; 137:378–87.

Brainard DH. The psychophysics toolbox. Spat Vis 1997; 10:433–6.

Brett M, Anton JL, Valabregue R, Poline JB. Region of interest analysis using the MarsBar toolbox for SPM 99. Neuroimage 2002; 16:S497.

Brookwell ML, Bentall RP, Varese F. Externalizing biases and hallucinations in source-monitoring, self-monitoring and signal detection studies: a meta-analytic review. Psychol Med 2013; 43:2465–75.

Buckner RL, Andrews-Hanna JR, Schacter DL. The brain's default network: anatomy, function, and relevance to disease. Ann N Y Acad Sci 2008; 1124:1–38.

Carter CS, Heckers S, Nichols T, Pine DS, Strother S. Optimizing the design and analysis of clinical functional magnetic resonance imaging research studies. Biol Psychiatry 2008; 64:842–9.

Cauda F, Geminiani G, D'Agata F, Sacco K, Duca S, Bagshaw AP, et al. Functional connectivity of the posteromedial cortex. PLoS One 2010; 5:e13107.

Clark A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav Brain Sci 2013; 36:181–204.

Corlett PR, Honey GD, Fletcher PC. Prediction error, ketamine and psychosis: an updated model. J Psychopharmacol 2016; 30:1145–55.

Corstens D, Longden E, McCarthy-Jones S, Waddingham R, Thomas N. Emerging perspectives from the hearing voices movement: implications for research and practice. Schizophr Bull 2014; 40:S285–94.

Daalman K, Boks MPM, Diederen KM, de Weijer Antoin D. , Blom Jan Dirk, Kahn RS, et al. The same or different? A phenomenological comparison of auditory verbal hallucinations in healthy and psychotic individuals. J Clin Psychiatry 2011; 72:320–5.

Daalman K, Verkooijen S, Derks EM, Aleman A, Sommer IEC. The influence of semantic top-down processing in auditory verbal hallucinations. Schizophr Res 2012; 139:82–6.

Davis MH, Johnsrude IS. Hierarchical processing in spoken language comprehension. J Neurosci Off J Soc Neurosci 2003; 23:3423–31.

Davis MH, Johnsrude IS. Hearing speech sounds: top-down influences on the interface between audition and speech perception. Hear Res 2007; 229:132–47.

Davis MH, Johnsrude IS, Hervais-Adelman A, Taylor K, McGettigan C. Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. J Exp Psychol Gen 2005; 134:222–41.

de Weijer AD, Neggers SFW, Diederen KMS, Mandl RCW, Kahn RS, Hulshoff Pol HE, et al. Aberrations in the arcuate fasciculus are associated with auditory verbal hallucinations in psychotic and in non-psychotic individuals. Hum Brain Mapp 2013; 34:626–34.

Dehaene-Lambertz G, Pallier C, Serniclaes W, Sprenger-Charolles L, Jobert A, Dehaene S. Neural correlates of switching from auditory to speech perception. Neuroimage 2005; 24:21–33.

Devinsky O, Morrell MJ, Vogt BA. Contributions of anterior cingulate cortex to behaviour. Brain J Neurol 1995; 118 (Pt 1):279–306.

Diederen KMJ, Daalman K, de Weijer AD, Neggers SFW, van Gastel W, Blom JD, et al. Auditory hallucinations elicit similar brain activation in psychotic and nonpsychotic individuals. Schizophr Bull 2012; 38:1074–82.

Diederen KMJ, De Weijer AD, Daalman K, Blom JD, Neggers SFW, Kahn RS, et al. Decreased language lateralization is characteristic of psychosis, not auditory hallucinations. Brain 2010; 133:373–44.

Du Y, Buchsbaum BR, Grady CL, Alain C. Noise differentially impacts phoneme representations in the auditory and speech motor systems. Proc Natl Acad Sci USA 2014; 111:7126–31.

Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, et al. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage 2005; 25:1325–35.

Eisner F, McGettigan C, Faulkner A, Rosen S, Scott SK. Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. J Neurosci 2010; 30:7179–86.

Fernyhough C, Bland K, Meins E, Coltheart M. Imaginary companions and young children's responses to ambiguous auditory stimuli: implications for typical and atypical development. J Child Psychol Psychiatry 2007; 48:1094–101.

Fletcher PC, Frith C. Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. Nat Rev Neurosci 2009; 10:48–58.

Friston KJ. Hallucinations and perceptual inference. Behav Brain Sci 2005; 28:764–6.

Gagnepain P, Henson RN, Davis MH. Temporal predictive codes for spoken words in auditory cortex. Curr Biol CB 2012; 22:615–21.

Grossberg S. How hallucinations may arise from brain mechanisms of learning, attention, and volition. J Int Neuropsychol Soc 2000; 6:583–92.

Haddock G, McCarron J, Tarrier N, Faragher E. Scales to measure dimensions of hallucinations and delusions: the psychotic symptom rating scales (PSYRATS). Psychol Med 1999; 29:879–89.

Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, et al. 'Sparse' temporal sampling in auditory fMRI. Hum Brain Mapp 1999; 7:213–23.

Hohwy J. The predictive mind. Oxford, UK: Oxford University Press; 2014.

Hunter MD, Eickhoff SB, Miller TWR, Farrow TFD, Wilkinson ID, Woodruff PWR. Neural activity in speech-sensitive auditory cortex during silence. Proc Natl Acad Sci USA 2006; 103:189–94.

Jardri R, Denève S. Circular inferences in schizophrenia. Brain 2013a; 136:3227–42.

Jardri R, Denève S. Computational models of hallucinations. In: Jardri R, Cachia A, Thomas P, Pins D, editors. The neuroscience of hallucinations. New York, NY: Springer; 2013b. p. 289–313.

Jardri R, Duverne S, Litvinova AS, Denève S. Experimental evidence for circular inference in schizophrenia. Nat Commun 2017; 8:14218.

Jardri R, Hugdahl K, Hughes M, Brunelin J, Waters F, Alderson-Day B, et al. Are hallucinations due to an imbalance between excitatory and inhibitory influences on the brain? Schizophr Bull 2016; 42:1124–34.

Jardri R, Pouchet A, Pins D, Thomas P. Cortical activations during auditory verbal hallucinations in schizophrenia: a coordinate-based meta-analysis. Am J Psychiatry 2011; 168:73–81.

Jardri R, Thomas P, Delmaire C, Delion P, Pins D. The neurodynamic organization of modality-dependent hallucinations. Cereb Cortex 2013; 23:1108–17.

Jarosz A, Wiley J. What are the odds? A practical guide to computing and reporting Bayes Factors. J Probl Solving 2014; 7:2–9.

JASP Team (2017). JASP (Version 0.8.1.2). [Computer Software].

Johns LC, Kompus K, Connell M, Humpston C, Lincoln TM, Longden E, et al. Auditory verbal hallucinations in persons with and without a need for care. Schizophr Bull 2014; 40:255–64.

Johns LC, Nazroo JY, Bebbington P, Kuipers E. Occurrence of hallucinations in a community sample. Schizophr Res 1998; 29:23.

Johns LC, van Os J. The continuity of psychotic experiences in the general population. Clin Psychol Rev 2001; 21:1125–41.

Kompus K, Falkenberg LE, Bless JJ, Johnsen E, Kroken RA, Kråkvik B, et al. The role of the primary auditory cortex in the neural mechanism of auditory verbal hallucinations. Front Hum Neurosci 2013; 7:144.

Kråkvik B, Larøi F, Kalhovde AM, Hugdahl K, Kompus K, Salvesen Ø, et al. Prevalence of auditory verbal hallucinations in a general population: A group comparison study. Scand J Psychol 2015; 56:508–15.

Kühn S, Gallinat J. Quantitative meta-analysis on state and trait aspects of auditory verbal hallucinations in schizophrenia. Schizophr Bull 2012; 38:779–86.

Langner R, Eickhoff SB. Sustaining attention to simple tasks: a meta-analytic review of the neural mechanisms of vigilant attention. Psychol Bull 2013; 139:870–900.

Leech R, Kamourieh S, Beckmann CF, Sharp DJ. Fractionating the default mode network: distinct contributions of the ventral and dorsal posterior cingulate cortex to cognitive control. J Neurosci 2011; 31:3217–24.

Leech R, Sharp DJ. The role of the posterior cingulate cortex in cognition and disease. Brain J Neurol 2014; 137:12–32.

Leptourgos P, Notredame CE, Jardri R, Denève S. Perceptual inferences in schizophrenia: a preliminary study on healthy participants. Eur Psychiatry 2015; 30:S113–14.

Lewis-Hanna LL, Hunter MD, Farrow TFD, Wilkinson ID, Woodruff PWR. Enhanced cortical effects of auditory stimulation and auditory attention in healthy individuals prone to auditory hallucinations during partial wakefulness. Neuroimage 2011; 57:1154–61.

Lima CF, Krishnan S, Scott SK. Roles of supplementary motor areas in auditory processing and auditory imagery. Trends Neurosci 2016; 39:527–42.

Linden DEJ, Thornton K, Kuswanto CN, Johnston SJ, van de Ven V, Jackson MC. The brain's voices: comparing nonclinical auditory hallucinations and imagery. Cereb Cortex 2011; 21:330–7.

Mayka MA, Corcos DM, Leurgans SE, Vaillancourt DE. Three-dimensional locations and boundaries of motor and premotor cortices as defined by functional brain imaging: a meta-analysis. Neuroimage 2006; 31:1453–74.

McGettigan C, Evans S, Rosen S, Agnew ZK, Shah P, Scott SK. An application of univariate and multivariate approaches in fMRI to quantifying the hemispheric lateralization of acoustic and linguistic processes. J Cogn Neurosci 2012; 24:636–52.

McGuire PK, Silbersweig DA, Murray RM, David AS, Frackowiak RS, Frith CD. Functional anatomy of inner speech and auditory verbal imagery. Psychol Med 1996; 26:29–38.

Mooney CM. Age in the development of closure ability in children. Can J Psychol Can Psychol 1957; 11:219–26.

Morrison AP, Wells A, Nothard S. Cognitive factors in predisposition to auditory and visual hallucinations. Br J Clin Psychol 2000; 39:67–78.

Möttönen R, Calvert GA, Jääskeläinen IP, Matthews PM, Thesen T, Tuomainen J, et al. Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. Neuroimage 2006; 30:563–9.

Nelson HE. National adult reading test. Test manual. Windsor: NFER-Nelson; 1982.

Obleser J, Kotz SA. Expectancy constraints in degraded speech modulate the language comprehension network. Cereb Cortex 2010; 20:633–40.

Orr JM, Weissman DH. Anterior cingulate cortex makes 2 contributions to minimizing distraction. Cereb Cortex 2009; 19:703–11.

Peters E, Ward T, Jackson M, Morgan C, Charalambides M, McGuire P, et al. Clinical, socio-demographic and psychological characteristics in individuals with persistent psychotic experiences with and without a 'need for care'. World Psychiatry 2016; 15:41–52.

Poldrack RA, Baker CI, Durnez J, Gorgolewski KJ, Matthews PM, Munafò MR, et al. Scanning the horizon: towards transparent and reproducible neuroimaging research. Nat Rev Neurosci 2017; 18:115–26.

Powers AR, Kelley M, Corlett PR. Hallucinations as top-down effects on perception. Biol Psychiatry Cogn Neurosci Neuroimaging 2016; 1:393–400.

Powers AR, Kelley MS, Corlett PR. Varieties of voice-hearing: psychics and the psychosis continuum. Schizophr Bull 2017; 43:84–98.

Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL. A default mode of brain function. Proc Natl Acad Sci USA 2001; 98:676–82.

Raij TT, Riekki TJJ. Poor supplementary motor area activation differentiates auditory verbal hallucination from imagining the hallucination. Neuroimage Clin 2012; 1:75–80.

Rao RP, Ballard DH. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 1999; 2:79–87.

Remez RE, Dubowski KR, Davids ML, Thomas EF, Paddu NU, Grossman YS, et al. Estimating speech spectra for copy synthesis by linear prediction and by hand. J Acoust Soc Am 2011; 130:2173–8.

Remez RE, Rubin PE, Pisoni DB, Carrell TD. Speech perception without traditional speech cues. Science 1981; 212:947–9.

Romme MA, Escher AD. Hearing voices. Schizophr Bull 1989; 15:209–16.

Rosen S, Wise RJS, Chadha S, Conway EJ, Scott SK. Hemispheric asymmetries in speech perception: sense, nonsense and modulations. PLoS One 2011; 6:e24672.

Rouder JN, Morey RD, Verhagen J, Swagman AR, Wagenmakers EJ. Bayesian analysis of factorial designs. Psychol Methods 2017; 22:304–21.

Schulz GM, Varga M, Jeffries K, Ludlow CL, Braun AR. Functional neuroanatomy of human vocalization: An H215O PET study. Cereb Cortex 2005; 15:1835–47.

Shanmugalingam P, McGettigan C, Agnew Z, Rosen S, Scott SK. Sinewave speech and non-speech can reveal distinct speech-mode and speech-intelligibility effects in the cortical speech perception network. New Orleans, LA: Society for Neuroscience; 2012.

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. Science 1995; 270:303–4.

Shergill SS, Brammer MJ, Williams SCR, Murray RM, McGuire PK. Mapping auditory hallucinations in schizophrenia using functional magnetic resonance imaging. Arch Gen Psychiatry 2000; 57:1033.

Simons CJP, Tracy DK, Sanghera KK, O'Daly O, Gilleen J, Dominguez MG, et al. Functional magnetic resonance imaging of inner speech in schizophrenia. Biol Psychiatry 2010; 67:232–7.

Sohoglu E, Peelle JE, Carlyon RP, Davis MH. Predictive top-down integration of prior knowledge during speech perception. J Neurosci 2012; 32:8443–53.

Sommer IE, Daalman K, Rietkerk T, Diederen KMJ, Bakker S, Wijkstra J, et al. Healthy individuals with auditory verbal hallucinations; who are they? Psychiatric assessments of a selected sample of 103 subjects. Schizophr Bull 2010; 36:633–41.

Szechtman H, Woody E, Bowers KS, Nahmias C. Where the imaginal appears real: a positron emission tomography study of auditory hallucinations. Proc Natl Acad Sci USA 1998; 95:1956–60.

Teufel C, Subramaniam N, Dobler V, Perez J, Finnemann J, Mehta PR, et al. Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. Proc Natl Acad Sci USA 2015; 112:13401–6.

Thomas N, Rossell SL, Waters F. The changing face of hallucination research: the International Consortium on Hallucination Research (ICHR) 2015 meeting report. Schizophr Bull 2016; 42:891–5.

van Lutterveld R, Diederen KMJ, Otte WM, Sommer IE. Network analysis of auditory hallucinations in nonpsychotic individuals. Hum Brain Mapp 2014; 35:1436–45.

Varese F, Barkus E, Bentall R. Dissociation mediates the relationship between childhood trauma and hallucination-proneness. Psychol Med 2012; 42:1025.

Vercammen A, Aleman A. Semantic expectations can induce false perceptions in hallucination-prone individuals. Schizophr Bull 2010; 36:151–6.

Vercammen A, de Haan EHF, Aleman A. Hearing a voice in the noise: auditory hallucinations and speech perception. Psychol Med 2008; 38:1177–84.

Vogt BA. Midcingulate cortex: structure, connections, homologies, functions and diseases. J Chem Neuroanat 2016; 74:28–46.

Vouloumanos A, Kiehl KA, Werker JF, Liddle PF. Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. J Cogn Neurosci 2001; 13:994–1005.

Zmigrod L, Garrison JR, Carr J, Simons JS. The neural mechanisms of hallucinations: a quantitative meta-analysis of neuroimaging studies. Neurosci Biobehav Rev 2016; 69:113–23.