1

2

Distinct Processing of Ambiguous Speech in People with Non-Clinical Auditory Verbal Hallucinations

5

6

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†].

9

10

[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.

21

*Indicates joint first authors

†Indicates joint last authors

24

25

Corresponding author: Dr. Ben Alderson-Day, Department of Psychology, Durham University, Science Laboratories, South Road, Durham, DH1 3LE, UK. Tel: +441913348147. Email: benjamin.alderson-day@durham.ac.uk

29

Short title: Speech processing and voice hallucinations

31

Keywords: hallucination, psychosis, auditory system, schizophrenia, imaging.

33

Abstract

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 away 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 an fMRI scan while passively listening to degraded speech ('sine-wave' speech, SWS), 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 SWS 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 behavioral 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   Introduction

2

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

18

19   Concerns about stigma make the recruitment of non-clinical voice-hearers extremely
20   challenging: consequently, only a handful of studies have sought to examine the
21   neurocognitive features of NCVH (e.g. Linden *et al.*, 2011; Kompus *et al.*, 2013). The
22   most successful of these was conducted in Utrecht, Holland, which initially identified
23   103 people with frequent NCVH who did not qualify for a psychiatric diagnosis
24   (Sommer *et al.*, 2010). To date, this remains the only project to have managed to run
25   neuroimaging studies in NCVH samples greater than 10 (Diederen *et al.*, 2012; de
26   Weijer *et al.*, 2013; van Lutterveld *et al.*, 2014). These studies have shown that when
27   hearing voices, people with NCVH and clinical AVH engage similar brain networks
28   associated with speech and language processing, including the bilateral superior
29   temporal gyrus (STG), inferior frontal gyrus (IFG) and anterior insula (AI) (Diederen
30   *et al.*, 2012). The experience of NCVH likely also involves regions associated with
31   the generation and monitoring of speech-motor imagery, as well as sensorimotor
32   processes, such as the supplementary and pre-supplementary motor areas (SMA/pre-
33   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 AVH (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.

Sine-wave speech (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 their spontaneous processing of ambiguous sounds. NCVH participants and matched non-voice-hearing controls passively listened to intelligible and unintelligible SWS while being scanned in fMRI, 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 (see Fig.1). After 20 minutes 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).

[Figure 1 here]

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. Behaviorally, 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 behavioral 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 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, PAC) across groups.

Materials and Methods

Participants

The study included twelve NCVH participants and 17 non-voice-hearing control participants, matched for age, sex, handedness, education, and National Adult

Reading Test scores (Nelson, 1982) (see Tab.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.

NCVH 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; see 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, 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 an fMRI scanning session (see Supplementary Materials 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, Evans *et al.* (2012). Intelligible (intSWS) and unintelligible SWS (unintSWS) were identical to those previously used apart from being further noise-vocoded (Shannon *et al.*, 1995), a step we deliberately omitted in order to make them less noticeably speech-like. The only exception were the 'target' sounds, which were created by noise-vocoding a subset of 10 unintelligible SWS in order 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 Materials 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 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 (Sennheiser U.K., High Wycombe, Buckinghamshire, U.K.), and then played three more examples of target sounds along with five non-vocoded unintSWS 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).

fMRI task

Participants listened to the SWS sounds across two identical runs of 20 minutes, broken up into six 'blocks' that were marked with a visually presented text stimulus (Block 1, Block 2, *etc;* see Fig.1). Each run contained 45 intSWS trials, 45 unintSWS 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?

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 intSWS 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-minute resting-state scans before and after the passive listening run as part of a separate study.

Post-scan behavioral task

9

Following scanning, participants were played 50 SWS stimuli in a random order (25 intSWS, 25 unintSWS). 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 vs. non-speech), and β (bias in identifying speech as present or absent). The post-scanner task was self-paced and took approximately 15 minutes.

MRI acquisition

MRI scanning was completed on a 1.5T Siemens Avanto (Siemens AG, Erlangen, Germany) 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: TR = 8.4 s; acquisition time = 3.4s, TE = 0.5s, flip angle = 90°, 40 axial slices, 3mm^3 in plane resolution. For localization, high-resolution anatomical images were also acquired using a T1-weighted magnetization prepared rapid acquisition gradient echo sequence (MP-RAGE; TR = 2.73s, TE = 3.57ms, flip angle = 7°, 176 sagittal slices, voxel size = 1mm^3).

Auditory onsets occurred 5s (±0-1s 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, Basingstoke, UK) and MRI-compatible insert earphones (Sensimetrics Corporation, Malden, MA, USA). 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 T1 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 $2mm^3$ voxels, and smoothed using a Gaussian kernel of $8mm^3$ at full-width-half-maximum to ameliorate differences in intersubject localization. Responses for events of interest were modelled using a canonical hemodynamic response function. IntSWS, unintSWS, target sounds and visual stimuli (block titles) were modelled from their onsets with durations of 2 seconds, 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 (3 translations, 3 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 (intSWS, unintSWS, vigilance targets) against the implicit rest baseline. The following planned contrasts were also generated during first-level analyses:

    i) (intSWS run 1 + intSWS run 2) - (unintSWS run 1 + unintSWS 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) (intSWS run 2 - unintSWS run 2) - (unintSWS run 1 - intSWS run 1), corresponding to a larger intelligibility response on run 2 vs. 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) intSWS run 1 - unintSWS 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) intSWS run 2 - unintSWS 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 behavioral 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 behavioral performance. All statistical maps were thresholded at $p < .001$ peak-level uncorrected, cluster corrected with a family-wise error (FWE) at $p < .05$ 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 ROIs based on the full cluster extent of activated regions in the above analyses. Between-groups comparison of behavioral data was analyzed using two-tailed t-tests at $p <. 05$, unless otherwise specified.

Results

Behavioral Results

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 (*M:* 3.71 and 4.94 for NCVH and controls, respectively; $t[27] = -2.17$, $p = $

.039[1]). 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 (1 NCVH, 3 controls) 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 $p >$ .19). Participants with irregular button-press data were marked and checked for their influence on group comparisons of fMRI data (see below). Only one NCVH participant reported a 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 behavioral 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; see Supplementary Materials Tab.2).

fMRI Results

- Responses to intelligible and unintelligible SWS 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 (IPL), and posterior STG. No supra-threshold group differences were evident for either the combination of intelligible and unintelligible SWS vs. rest (i.e., the main effect of group during

---

[1] 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 = .041, Monte Carlo Method used with 2000 replications).

13

listening to sounds), nor any simple effects (i.e., the main effect of group during listening to intelligible-only SWS vs. rest and unintelligible-only SWS vs. rest).

[Figure 2 here]

- 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 medial part of the superior frontal gyrus (Tab. 2a and Figure 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 (Tab.2b 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 panel). 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 ($t[27] = 2.98$, $p = .006$), but the groups were similar for unintelligible SWS ($t[27] = -1.05$, $p = .301$). The reverse contrast (Controls > NCVH) yielded no significant clusters.

[Figure 3 here]

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, $p_{FWE} < .001$), indicating that greater

14

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} < .001; see Figure 3d). This change was specific to intelligible stimuli, i.e., no effect was evident for the change in responses to unintelligible stimuli (Figure 3d, right panel). 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 < .001), but responses to unintelligible SWS were similar across runs (t[28] = 0.12, p = .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 < .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 (see Tab.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} = .008) and in left middle frontal gyrus (MNI coordinates for peak voxel: -36, 54, 0, k = 190, t = 4.91, p_{FWE} = .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 < .001

1 uncorrected), which was consistent with the general enhancement of an intelligibility
2 effect across the whole scanning session.
3
4 *- Comparing responses in PAC*
5 The lack of supra-threshold group differences in responses to intelligible or
6 unintelligible stimuli over rest indicated that basic auditory processes were broadly
7 similar in controls and NCVH. To explore this further, we extracted average
8 responses to intelligible and unintelligible sounds in the bilateral primary auditory
9 cortices (defined as TE 1.0, 1.1, and 1.2 based on the SPM Anatomy Toolbox) and
10 conducted Bayesian inference testing on effects of group, intelligibility, and run. A
11 Bayesian mixed ANOVA was conducted using JASP (Love *et al.*, 2015), with the
12 default priors (Rouder *et al.*, 2016). When a model containing the group effect was
13 compared to one without it (i.e., the null hypothesis), the Bayes Factor (BF) was 0.54,
14 or 1:1.86 in favour of the null (in other words, the data were almost twice as likely to
15 occur under the null hypothesis). Evidence for any group-related interaction effects
16 was even weaker: BF values of 0.26, 0.26 and 0.57 were observed for models
17 containing group × run, group × intelligibility, and group × run × intelligibility,
18 respectively (i.e., 1:3.85, 1:3.85 and 1:1.75 in favour of the null)[2]. These values only
19 reflect anecdotal to substantial evidence in favor of the null hypothesis (Jarosz and
20 Wiley, 2014), but they nevertheless offer no evidence at all in favour of potential
21 group differences in PAC signal.
22
23 *- Individual differences in intelligibility responses*
24 To explore how early responders may have been identifying speech in the SWS, we
25 ran a whole-brain individual differences analysis, including recognition point as a
26 regressor in the Intelligible > Unintelligible SWS contrast. The intelligibility response
27 across runs 1 and 2 in left IFG was negatively related to the recognition point
28 (indicating that those who noticed speech earlier showed greater activation in these
29 regions; see Fig.4a and Tab.4). For run 1 only (i.e., before all participants were in
30 'speech mode'), the recognition point was negatively related to responses in the

---

[2] BF 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*., 2016).

1 middle cingulate cortex extending to parietal areas (Fig.4b) and positively related to

2 activation in medial prefrontal cortex (Fig.4c). We also ran the same analysis for an

3 index of voice-hearing in the NCVH participants (PSYRATS Physical Characteristics

4 from the past week; see Supplementary Materials); this indicated no significant

5 whole-brain correlations. However, a behavioral correlation was observed between

6 voice-hearing in the past week and recognition point (r = -.582, n = 12, p = .047),

7 such that a greater tendency to hear voices was associated with noticing speech earlier

8 in run 1 (Fig.1D). This correlation directly links auditory-perceptual processes, as

9 evaluated in the current study, with the magnitude of recent AVH.

10

11   [Figure 4 here]

12

13

14 **Discussion**

15

16 Despite decades of work on hallucinations, little is known about how they relate to

17 everyday perceptual mechanisms. Our research aimed to address this by studying the

18 interaction of expectation and perception in non-clinical voice-hearers. Knowledge

19 and expectations help us to interpret ambiguous signals in a range of contexts; in

20 some cases, this might lead to non-veridical sensations, but in other situations – such

21 as hearing sine-wave speech – such expectations might contribute to divining

22 meaningful signal from apparent noise (Davis and Johnsrude, 2007).

23

24 Behavioral evidence of NCVH hearing semantically congruent (but absent) speech in

25 white noise (Daalman *et al.*, 2012) and signal detection biases in people prone to

26 hallucinations (Brookwell *et al.*, 2013) has been used to argue for the existence of

27 attentional factors – such as expectation and prior knowledge – having a greater

28 influence on perception in people who hear voices. Our design, by initially disguising

29 the presence of speech from participants, allowed us to examine whether such an

30 influence can act spontaneously in NCVH, or requires the specific modulation of

31 expectation (in essence, a suggestibility effect). The subjective behavioral responses

32 of voice-hearers here – reporting the detection of speech content in the acoustics of

33 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. 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 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, 2013a), 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, 2013b; 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 (PAC), 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 (e.g. Obleser and Kotz, 2010) – also showed no group differences. Instead, differences were seen in a region of 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 AVH, via atypical modulation of sensory regions (see Allen *et al.*, 2007, for a review). 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 (e.g., 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 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 AVH (Linden *et al.*, 2011; 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, mPFC, and 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 mPFC, 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 (e.g. 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

1 networks responsible for cognitive control and external attention (Cauda *et al.*, 2010;
2 Leech *et al.*, 2011; Leech and Sharp, 2014).
3
4 Some limitations of the present study must be acknowledged. First, for practical
5 reasons – and because of the goals of the experiment – the behavioural assessment of
6 participants' ability to discriminate and understand SWS had to be conducted outside
7 the scanner and followed a long period of training and exposure to the stimuli. As
8 such, it is possible that any post-scan group differences were masked or trained out as
9 a result of the procedure, given that decoding of other kinds of degraded auditory
10 stimuli – such as noise-vocoded speech – can improve over time and with training
11 (Davis *et al.*, 2005). However, neither group performed at ceiling on the post-scan
12 task: keyword accuracy after scanning was reasonably low in both groups compared
13 to prior studies using distorted speech (McGettigan *et al.*, 2012), despite the fact that
14 speech/non-speech discrimination was good. In future studies it will be important to
15 assess NCVH participants' abilities to decode SWS under a variety of listening
16 conditions to measure decoding skill and adaptation more directly.
17
18 Second, we are reliant on the accuracy of participants' self-reports to gauge when
19 participants noticed speech during run 1, and cannot know for sure what participants
20 were responding to when 'hearing' speech. Relying on self-report data is not
21 uncommon in hallucinations research and retrospective reporting of events in the
22 scanner has been used successfully to identify periods of voice-hearing (e.g. Jardri *et
23 al.*, 2013). Nevertheless, it is possible that NCVH participants were just more likely to
24 class any unusual stimuli as speech, rather than intelligible stimuli specifically. Two
25 pieces of evidence militate against such an interpretation, though: first, the lack of any
26 general group differences in the neural response to stimuli versus baseline (i.e., across
27 both intelligible and unintelligible SWS), and second, the lack of any evident speech
28 bias on the post-scan behavioral task. Notably, our brain data provides evidence in
29 favour of a selective effect for the discrimination of intelligible stimuli: an effect that
30 is hard to account by positing a non-specific response bias. Future studies could
31 further address the selectivity of the behavioral effect by testing whether differences
32 in recognition point also exist for a run without potentially intelligible SWS (this
33 would be evidence for a non-specific bias), or by assessing degraded speech
34 perception skills more comprehensively prior to training (e.g., Boebinger *et al.*, 2015).

1 Including such conditions in the current study would have compromised our ability to

2 test naïve participants' spontaneous responses to ambiguous stimuli.

3

4 Finally, we were restricted to a smaller sample of participants in the present study

5 than is generally recommended for clinical fMRI research (Carter *et al.*, 2008) and for

6 group comparisons in general fMRI studies (Poldrack *et al.*, 2017). Recruitment for

7 neuroimaging studies with NCVH groups is extremely challenging: the present

8 sample size is larger than other recent studies (Linden *et al.*, 2011; Kompus *et al.*,

9 2013), with the exception of the Utrecht cohort (e.g. Diederen et al., 2013). Prior

10 NCVH imaging studies have largely confined task-based fMRI investigations to

11 symptom capture (Linden et al., 2011, Diederen et al., 2012) or basic cognitive

12 paradigms, such as dichotic listening (Kompus et al., 2011) or verbal fluency

13 (Diederen et al., 2010), often with recourse to region-of-interest analysis and other

14 methods of constraining analysis (and statistical corrections) to selected brain regions.

15 To our knowledge, this is the first NCVH study to have successfully combined a

16 complex behavioral paradigm with imaging data to examine a potential mechanism

17 underlying hallucination, and while maintaining conservative whole-brain corrections.

18 Nevertheless, small sample sizes in neuroimaging research with clinical and non-

19 clinical voice-hearers is an enduring problem. As we have advocated elsewhere

20 (Alderson-Day *et al.*, 2016) the combination of fMRI data from multiple laboratories

21 provides one means of addressing this issue. The International Consortium of

22 Hallucinations Research (ICHR) is currently supporting ongoing mega-analytic

23 projects involving the combination of task-based, resting-state and structural MRI

24 data from people with AVH (Thomas *et al.*, 2016).

25

26 Notwithstanding the small sample size of the present study, it is also important to note

27 that the general response to intelligibility and general effects of training with SWS –

28 involved regions consistent with previous research on distorted speech. The primarily

29 left-lateralized network seen across both groups is consistent with intelligibility

30 effects using very similar stimuli (McGettigan *et al.*, 2012), as is the involvement of

31 the SMA (Rosen *et al.*, 2011). The involvement of left posterior STG seen

32 specifically following training also replicates prior findings using SWS (Möttönen *et*

33 *al.*, 2006). Thus, in general, these two groups of participants showed plausible

34 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 fMRI 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

2

3　*Interview session*

4　The voice phenomenology interview used to assess participants included questions

5　from the Psychotic Symptoms Ratings Scale (PSYRATS; Haddock *et al.*, 1999) and

6　the Positive and Negative Syndrome Scale (PANSS; Kay *et al.*, 1987). To allow for

7　indicative ratings of voice-hearing comparable to AVH, participants were specifically

8　scored on the auditory hallucinations subscale of the PSYRATS and positive and

9　negative subscales of the PANSS (PANSS-P and PANSS-N, respectively). The

10　Beliefs About Voices Questionnaire-Revised (BAVQ-R; Chadwick *et al.*, 2000) was

11　also originally included, although many participants preferred not to complete it

12　because of its focus on malevolent and dominant voices (which they deemed

13　irrelevant to their own experience).

14

15　Overall scores on these measures were similar to previous cohorts: positive and

16　negative scores on the PANSS were comparable to those reported by Linden et al.

17　(2011) (PANSS-P = 12, PANSS-N = 7), while individual PSYRATS item scores were

18　broadly in line with those reported by Daalman et al. (2011) for the Utrecht cohort .

19　The mean score of 4 for P3 (Hallucinations) indicated that participants' voices were

20　rated as 'Moderate' – occurring frequently but not continuously, with thinking and

21　behavior minimally affected (Kay *et al.*, 1987). As would be expected for a voice-

22　hearer group without other psychosis-like characteristics, i) ratings for positive

23　symptoms were significantly higher than negative symptoms ($t[11] = 10.86$, p <. 001)

24　and ii) ratings for hallucinations were significantly higher than delusions ($t[11] =$

25　5.86, p < .001). On average, voices occurred around once a day, for seconds at a time,

26　were located inside the head or close by, and contained very little negative or

27　distressing content. No participants reported that their voices were problematic or

28　disruptive to their everyday lives and all were either in work, education, or retired.

29

30　As in some cases a number of weeks passed between the full interview and

31　participants' scanning session (*M(SD)* = 77.64 (55.28) days), items 1-4 of the

32　PSYRATS (the 'physical characteristics' subscale, assessing frequency, duration,

33　location, and volume of voices) were re-administered via a short phone interview in

the week of the scan. There was no difference observed between scores at interview

or in the week of the scan ($t(11) = 1.69$, $p = .12$). As the most recent index of voice-

hearing, this measure was then used for correlation analysis with the behavioural and

neuroimaging results from the scanning session.

Supplementary Table 1. PSYRATS item ratings for voice-hearing characteristics, compared with the 'Utrecht' sample (Daalman *et al.*, 2011)

	Item	Present Study		Daalman et al. (2011)	
		M	*SD*	*M*	*SD*
Frequency	1	2.08	1.00	3.53	1.26
Duration	2	1.75	1.14	1.53	0.73
Location	3	2.00	0.85	2.21	1.15
Volume	4	2.00	1.13	1.81	0.65
Beliefs About Voice					
Origin	5	2.25	0.87	3.17	1.13
Emotional Valence	6-8	1.42	1.62	1.69	3.05
Total Distress	9-10	1.66	1.07	0.63	1.33
Control	11	1.17	1.19	1.77	1.49

Note. PSYRATS = Psychotic Symptoms Rating Scale. Individual items are scored

from 0-4 by the interviewer. E.g. Frequency: 0 – Voices not present, 1 – Once a week,

2 – Once a day, 3 – Once an hour, 4 – Continuously.

Sine-wave speech stimulus preparation

Potentially intelligible SWS sentences (intSWS) were created by applying the

frequency and amplitude estimates of the formants from the same original sentence,

while unintelligible SWS control stimuli (unintSWS) combined spectral and

amplitude tracks from different sentences (int$S_{mod}A_{mod}$ and $S_{mod}A_{mod}$ respectively in

Rosen et al., 2011). For noise-vocoded target sounds, the stimulus waveform was

passed through a bank of 16 analysis filters (sixth-order Butterworth) with frequency

responses crossing 3 dB down from the passband peak. Envelope extraction at the

output of each analysis filter was carried out using full-wave rectification and second-

order Butterworth low-pass filtering at 60 Hz. The envelopes were then multiplied by

a white noise, and each filtered by a sixth-order Butterworth IIR output filter identical

1   to the analysis filter. The root-mean-square (rms) level from each output filter was set

2   to be equal to the rms level of the original analysis outputs. Finally, the modulated

3   outputs were summed together. The cross-over frequencies for both filter banks (over

4   the frequency range of 70–5000 Hz) were calculated using an equation relating

5   position on the basilar membrane to its best frequency (Greenwood, 1990). Noise-

6   vocoding of the SWS increased the bandwidth of the sinewaves changing their timbre,

7   and excitation with noise ensured that they had a 'noisier/rougher' quality compared to

8   the intelligible and unintelligible SWS stimuli.

9

10   **Supplementary Table 2.** Performance on the post-scan behavioral task

	NCVH		Control		
	M	*SD*	*M*	*SD*	*p*
Keyword Accuracy	38.33%	14.04%	45.41%	22.72%	0.348
d' (sensitivity)	1.95	0.73	2.02	0.57	0.791
β (bias)	0.68	0.45	0.58	0.85	0.712

11   Note. NCVH = non-clinical voice-hearers.

12

13

14

15

16

17

References

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

Alderson-Day B. Do you hear voices? You are not alone. 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: 1110–1123.

Allen P, Aleman A, Mcguire PK. Inner speech models of auditory verbal hallucinations: Evidence from behavioural and neuroimaging studies. Int. Rev. Psychiatry 2007; 19: 407–415.

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: 125–141.

Beavan V, Read J, Cartwright C. The prevalence of voice-hearers in the general population: a literature review. J Ment Health 2011; 20: 281–292.

Bench J, Kowal Å, Bamford J. The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. Br. J. Audiol. 1979; 13: 108–112.

Benson RR, Richardson M, Whalen DH, Lai S. Phonetic processing areas revealed by sinewave speech and acoustically similar non-speech. NeuroImage 2006; 31: 342–353.

Bentall RP, Slade PD. Reliability of a scale measuring disposition towards hallucination: a brief report. Personal. Individ. Differ. 1985; 6: 527–529.

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: 378–387.

Brainard DH. The psychophysics toolbox. Spat. Vis. 1997; 10: 433–436.

Brett M, Anton J-L, Valabregue R, Poline J-B. 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: 2465–2475.

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.

1   Carter CS, Heckers S, Nichols T, Pine DS, Strother S. Optimizing the design and
2   analysis of clinical functional magnetic resonance imaging research studies. Biol.
3   Psychiatry 2008; 64: 842–849.

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

6   Chadwick P, Lees S, Birchwood M. The revised Beliefs About Voices Questionnaire
7   (BAVQ-R). Br. J. Psychiatry 2000; 177: 229–232.

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

10   Corlett PR, Honey GD, Fletcher PC. Prediction error, ketamine and psychosis: An
11   updated model. J. Psychopharmacol. (Oxf.) 2016; 30: 1145–1155.

12   Corstens D, Longden E, McCarthy-Jones S, Waddingham R, Thomas N. Emerging
13   perspectives from the Hearing Voices Movement: Implications for research and
14   practice. Schizophr. Bull. 2014; 40: S285–S294.

15   Daalman K, Boks MPM, Diederen KM, Weijer, Antoin D. de, Blom, Jan Dirk, Kahn
16   RS, et al. The same or different? A phenomenological comparison of auditory verbal
17   hallucinations in healthy and psychotic individuals. J. Clin. Psychiatry 2011; 72: 320–
18   325.

19   Daalman K, Verkooijen S, Derks EM, Aleman A, Sommer IEC. The influence of
20   semantic top-down processing in auditory verbal hallucinations. Schizophr. Res.
21   2012; 139: 82–86.

22   Davis MH, Johnsrude IS. Hierarchical processing in spoken language comprehension.
23   J. Neurosci. Off. J. Soc. Neurosci. 2003; 23: 3423–3431.

24   Davis MH, Johnsrude IS. Hearing speech sounds: Top-down influences on the
25   interface between audition and speech perception. Hear. Res. 2007; 229: 132–147.

26   Davis MH, Johnsrude IS, Hervais-Adelman A, Taylor K, McGettigan C. Lexical
27   information drives perceptual learning of distorted speech: evidence from the
28   comprehension of noise-vocoded sentences. J. Exp. Psychol. Gen. 2005; 134: 222–
29   241.

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

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

35   Diederen KMJ, Daalman K, de Weijer AD, Neggers SFW, van Gastel W, Blom JD, et
36   al. Auditory hallucinations elicit similar brain activation in psychotic and
37   nonpsychotic individuals. Schizophr. Bull. 2012; 38: 1074–1082.

1   Du Y, Buchsbaum BR, Grady CL, Alain C. Noise differentially impacts phoneme
2   representations in the auditory and speech motor systems. Proc. Natl. Acad. Sci. 2014;
3   111: 7126–7131.

4   Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, et al. A new
5   SPM toolbox for combining probabilistic cytoarchitectonic maps and functional
6   imaging data. NeuroImage 2005; 25: 1325–1335.

7   Eisner F, McGettigan C, Faulkner A, Rosen S, Scott SK. Inferior frontal gyrus
8   activation predicts individual differences in perceptual learning of cochlear-implant
9   simulations. J. Neurosci. 2010; 30: 7179–7186.

10   Fernyhough C, Bland K, Meins E, Coltheart M. Imaginary companions and young
11   children's responses to ambiguous auditory stimuli: implications for typical and
12   atypical development. J. Child Psychol. Psychiatry 2007; 48: 1094–1101.

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

15   Friston KJ. Hallucinations and perceptual inference. Behav. Brain Sci. 2005; 28: 764–
16   766.

17   Gagnepain P, Henson RN, Davis MH. Temporal predictive codes for spoken words in
18   auditory cortex. Curr. Biol. CB 2012; 22: 615–621.

19   Greenwood DD. A cochlear frequency-position function for several species--29 years
20   later. J. Acoust. Soc. Am. 1990; 87: 2592–2605.

21   Grossberg S. How hallucinations may arise from brain mechanisms of learning,
22   attention, and volition. J. Int. Neuropsychol. Soc. 2000; 6: 583–592.

23   Haddock G, McCarron J, Tarrier N, Faragher E. Scales to measure dimensions of
24   hallucinations and delusions: the psychotic symptom rating scales (PSYRATS).
25   Psychol. Med. 1999; 29: 879–889.

26   Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, et al.
27   'Sparse' temporal sampling in auditory fMRI. Hum. Brain Mapp. 1999; 7: 213–223.

28   Hohwy J. The predictive mind. Oxford, United Kingdom: Oxford University Press;
29   2014.

30   Hunter MD, Eickhoff SB, Miller TWR, Farrow TFD, Wilkinson ID, Woodruff PWR.
31   Neural activity in speech-sensitive auditory cortex during silence. Proc. Natl. Acad.
32   Sci. 2006; 103: 189–194.

33   Jardri R, Denève S. Circular inferences in schizophrenia. Brain 2013a; 136: 3227–
34   3242.

35   Jardri R, Denève S. Computational models of hallucinations. In: Jardri R, Cachia A,
36   Thomas P, Pins D, editor(s). The neuroscience of hallucinations. Springer New York;
37   2013. b. 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: 1124–1134.

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: 1108–1117.

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

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: 255–264.

Johns LC, Nazroo JY, Bebbington P, Kuipers E. Occurrence of hallucinations in a community sample. Schizophr. Res. 1998; 1-2: 23.

Johns LC, van Os J. The continuity of psychotic experiences in the general population. Clin. Psychol. Rev. 2001; 21: 1125–1141.

Kay SR, Fiszbein A, Opfer LA. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr. Bull. 1987; 13: 261.

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: 508–515.

Kühn S, Gallinat J. Quantitative meta-analysis on state and trait aspects of auditory verbal hallucinations in schizophrenia. Schizophr. Bull. 2012; 38: 779–786.

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: 3217–3224.

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

1   Leptourgos P, Notredame CE, Jardri R, Denève S. Perceptual inferences in
2   schizophrenia: A preliminary study on healthy participants. Eur. Psychiatry 2015; 30:
3   S113–S114.

4   Lewis-Hanna LL, Hunter MD, Farrow TFD, Wilkinson ID, Woodruff PWR.
5   Enhanced cortical effects of auditory stimulation and auditory attention in healthy
6   individuals prone to auditory hallucinations during partial wakefulness. NeuroImage
7   2011; 57: 1154–1161.

8   Lima CF, Krishnan S, Scott SK. Roles of supplementary motor areas in auditory
9   processing and auditory imagery. Trends Neurosci. 2016; 39: 527–542.

10 Linden DEJ, Thornton K, Kuswanto CN, Johnston SJ, Ven V van de, Jackson MC.
11 The brain's voices: Comparing nonclinical auditory hallucinations and imagery.
12 Cereb. Cortex 2011; 21: 330–337.

13 Love J, Selker R, Marsman M, Jamil T, Dropmann D, Verhagen A, et al. JASP
14 (Version 0.7)[Computer software]. 2015

15 van Lutterveld R, Diederen KMJ, Otte WM, Sommer IE. Network analysis of
16 auditory hallucinations in nonpsychotic individuals. Hum. Brain Mapp. 2014; 35:
17 1436–1445.

18 Mayka MA, Corcos DM, Leurgans SE, Vaillancourt DE. Three-dimensional locations
19 and boundaries of motor and premotor cortices as defined by functional brain
20 imaging: A meta-analysis. NeuroImage 2006; 31: 1453–1474.

21 McGettigan C, Evans S, Rosen S, Agnew ZK, Shah P, Scott SK. An application of
22 univariate and multivariate approaches in fMRI to quantifying the hemispheric
23 lateralization of acoustic and linguistic processes. J. Cogn. Neurosci. 2012; 24: 636–
24 652.

25 McGuire PK, Silbersweig DA, Murray RM, David AS, Frackowiak RS, Frith CD.
26 Functional anatomy of inner speech and auditory verbal imagery. Psychol. Med.
27 1996; 26: 29–38.

28 Mooney CM. Age in the development of closure ability in children. Can. J. Psychol.
29 Can. Psychol. 1957; 11: 219–226.

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

32 Möttönen R, Calvert GA, Jääskeläinen IP, Matthews PM, Thesen T, Tuomainen J, et
33 al. Perceiving identical sounds as speech or non-speech modulates activity in the left
34 posterior superior temporal sulcus. NeuroImage 2006; 30: 563–569.

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

36 Obleser J, Kotz SA. Expectancy constraints in degraded speech modulate the
37 language comprehension network. Cereb. Cortex 2010; 20: 633–640.

Orr JM, Weissman DH. Anterior cingulate cortex makes 2 contributions to minimizing distraction. Cereb. Cortex 2009; 19: 703–711.

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: 115–126.

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. 2001; 98: 676–682.

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: 2173–2178.

Remez RE, Rubin PE, Pisoni DB, Carrell TD. Speech perception without traditional speech cues. Science 1981; 212: 947–949.

Romme MA, Escher AD. Hearing voices. Schizophr. Bull. 1989; 15: 209–216.

Rosen S, Wise RJS, Chadha S, Conway E-J, 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 E-J. Bayesian analysis of factorial designs. Psychol. Methods 2016: Online First.

Schulz GM, Varga M, Jeffires K, Ludlow CL, Braun AR. Functional neuroanatomy of human vocalization: An H215O PET study. Cereb. Cortex 2005; 15: 1835–1847.

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. Society for Neuroscience: New Orleans, LA: 2012.
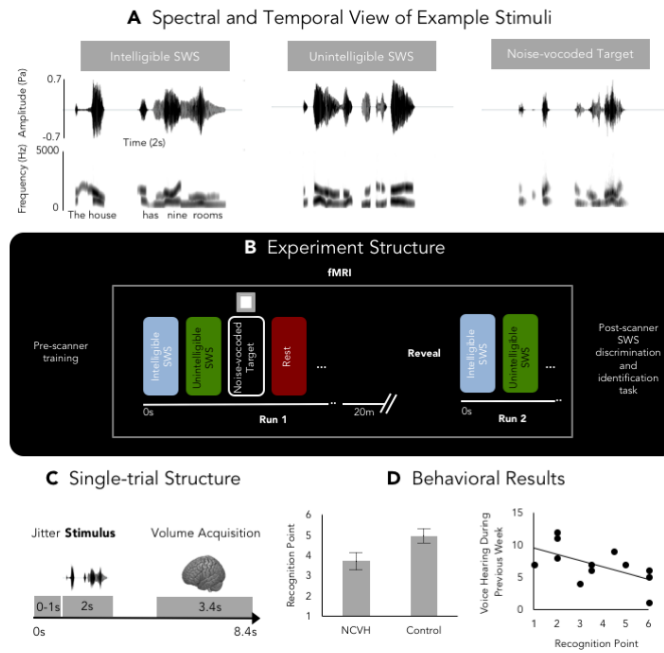
1 Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with
2 primarily temporal cues. Science 1995; 270: 303–304.

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

6 Simons CJP, Tracy DK, Sanghera KK, O'Daly O, Gilleen J, Dominguez M-G, et al.
7 Functional magnetic resonance imaging of inner speech in schizophrenia. Biol.
8 Psychiatry 2010; 67: 232–237.

9 Sohoglu E, Peelle JE, Carlyon RP, Davis MH. Predictive top-down integration of
10 prior knowledge during speech perception. J. Neurosci. 2012; 32: 8443–8453.

11 Sommer IE, Daalman K, Rietkerk T, Diederen KMJ, Bakker S, Wijkstra J, et al.
12 Healthy individuals with auditory verbal hallucinations; who are they? Psychiatric
13 assessments of a selected sample of 103 subjects. Schizophr. Bull. 2010; 36: 633–641.

14 Szechtman H, Woody E, Bowers KS, Nahmias C. Where the imaginal appears real: A
15 positron emission tomography study of auditory hallucinations. Proc. Natl. Acad. Sci.
16 U. S. A. 1998; 95: 1956–1960.

17 Teufel C, Subramaniam N, Dobler V, Perez J, Finnemann J, Mehta PR, et al. Shift
18 toward prior knowledge confers a perceptual advantage in early psychosis and
19 psychosis-prone healthy individuals. Proc. Natl. Acad. Sci. 2015; 112: 13401–13406.

20 Thomas N, Rossell SL, Waters F. The changing face of hallucination research: The
21 International Consortium on Hallucination Research (ICHR) 2015 meeting report.
22 Schizophr. Bull. 2016; 42: 891–895.

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

25 Vercammen A, Aleman A. Semantic expectations can induce false perceptions in
26 hallucination-prone individuals. Schizophr. Bull. 2010; 36: 151–156.

27 Vercammen A, de Haan EHF, Aleman A. Hearing a voice in the noise: auditory
28 hallucinations and speech perception. Psychol. Med. 2008; 38: 1177–1184.

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

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

34 de Weijer AD, Neggers SFW, Diederen KMS, Mandl RCW, Kahn RS, Hulshoff Pol
35 HE, et al. Aberrations in the arcuate fasciculus are associated with auditory verbal
36 hallucinations in psychotic and in non-psychotic individuals. Hum. Brain Mapp.
37 2013; 34: 626–634.

1   Zmigrod L, Garrison JR, Carr J, Simons JS. The neural mechanisms of hallucinations:
2   A quantitative meta-analysis of neuroimaging studies. Neurosci. Biobehav. Rev.
3   2016; 69: 113–123.

4

5

6

A Spectral and Temporal View of Example Stimuli

Intelligible SWS | Unintelligible SWS | Noise-vocoded Target

B Experiment Structure

C Single-trial Structure   **D** Behavioral Results

1

2

3 **Figure 1.** Participants were scanned in fMRI while (A) listening to intelligible SWS,

4 unintelligible SWS, or noise-vocoded, unintelligible target sounds; (B) listening and

5 rest trials were presented in a pseudo-random order across two 20-minute runs,

6 divided by a 'reveal' period including training to understand SWS stimuli; (C) each

7 trial lasted 8.4s, including jitter, a 2s stimulus and 3.4s of volume acquisition; (D)

8 NCVH participants recognized speech being present earlier than control participants

9 during run 1 (left panel), and this correlated with voice-hearing during the previous

10 week (PSYRATS – Physical Characteristics subscale). NCVH = non-clinical voice-

11 hearing; PSYRATS = Psychotic Symptoms Rating Scale; SWS = sine-wave speech.

12
13
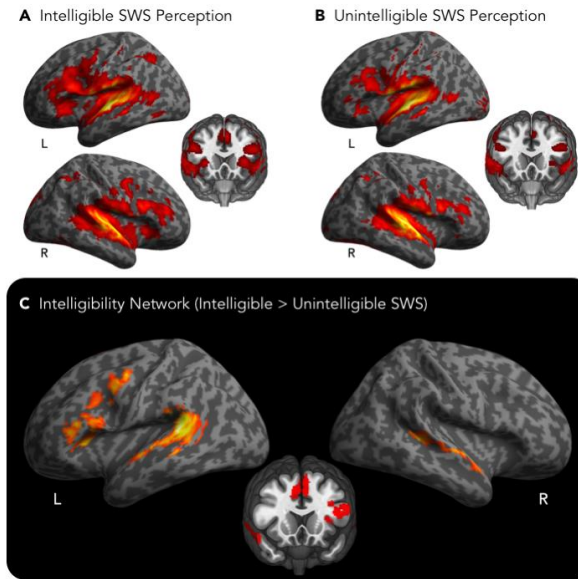
Figure 2. Responses vs. rest baseline to intelligible SWS (A), unintelligible SWS (B), and the difference between them, i.e. the intelligibility effect (C). SWS = sine-wave speech. Activation maps are presented at an uncorrected threshold of $p < .001$ peak level, FWE corrected ($p < .05$) at cluster level.
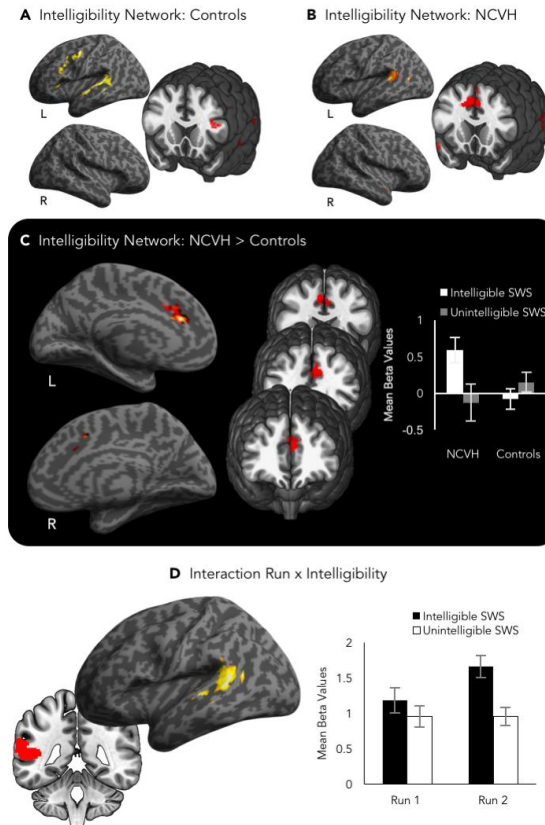
Figure 3. Intelligibility responses in control participants (A), in voice-hearers (B), between-group differences in the intelligibility effect (C), and the change in the intelligibility effect following training with intelligible SWS, both groups combined (D). Beta values shown in (C) are extracted from 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 coordinates for peak voxel: -50, -48, 10) identified in the Run × Intelligibility whole-brain interaction. SWS = sine-wave speech; NCVH = non-clinical voice-hearers. Activation maps are presented at an uncorrected threshold of $p < .001$ peak level, FWE corrected ($p < .05$) at cluster level.

Correlations Between When Participants Noticed Words and Activity in Intelligible > Unintelligible SWS

1

Figure 4. Correlations between the recognition point when participants noticed words and intelligibility response across both runs (A), and in run 1 only (B and C). SWS = sine-wave speech, NCVH = non-clinical voice-hearing. Activation maps are presented at an uncorrected threshold of p < .001 peak level, FWE corrected (p < .05) at cluster level.

7
8

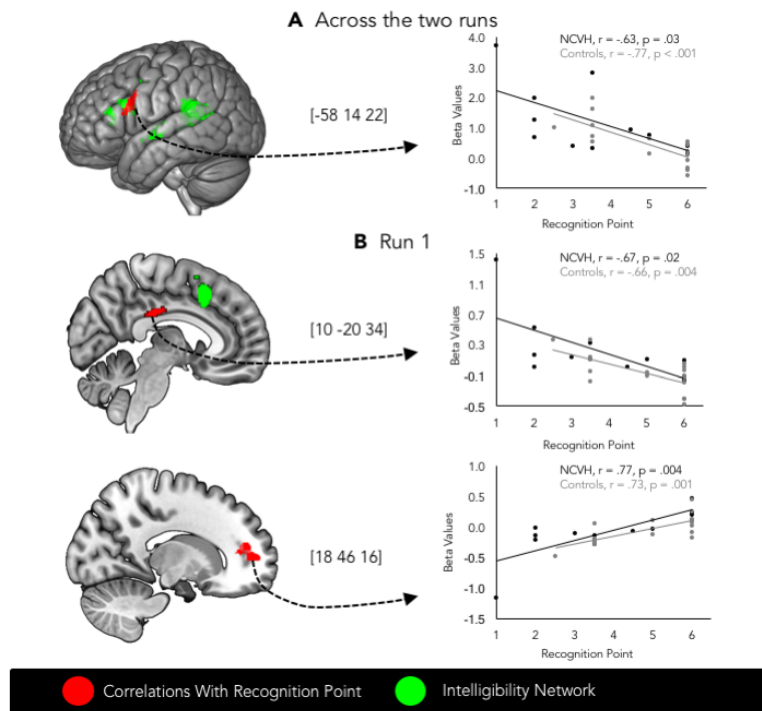**Table 1.** Participant demographic and clinical characteristics

	NCVH		Control		*p*
Sex	8F/4M		12F/5M		0.822
Handedness	11R/1L		14R/3L		0.474
	M	*SD*	*M*	*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	-	-	-

Note. F = female; M = male; R = right; L = left; NCVH = Non-Clinical Voice-Hearers; NART = National Adult Reading Test; PSYRATS-AH = Psychotic Symptoms Rating Scale - Auditory Hallucinations; PANSS = Positive & 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.

1   Table 2. Responses to Intelligible > Unintelligible SWS, across both runs and groups

2   combined (A), and group differences in intelligibility responses (B).

Location	x	y	z	# Voxels	t	z	p_{FWE}
A. Runs and groups combined							
L Superior Temporal Gyrus	-62	-42	16	1763	8.24	5.78	< .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	< .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	< .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	< .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	.001
B. Group differences							
L Anterior Cingulate Cortex	-4	34	26	539	4.81	4.05	< .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

1 Note. These results are presented at an uncorrected threshold of p < .001 peak level, FWE

2 corrected (p < .05) at cluster level. L = Left; R = Right. We report a maximum of 15 grey

3 matter local maxima (that are more than 8 mm apart) per cluster.

4

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

run

Contrast	Location	x	y	z	# Voxels	t	z	p_{FWE}
Run 1	-	-	-	-	-	-	-	-
Run 2	L Inferior Parietal Cortex	-68	-42	20	2125	8.96	6.05	< .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	< .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	< .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	< .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	.019
R Inferior Frontal Gyrus	36	32	-6		4.22	3.74	

1 Note. These results are presented at an uncorrected threshold of p < .001 peak level,

2 FWE corrected (p < .05) at cluster level. L = Left; R = Right. We report a maximum

3 of 15 grey matter local maxima (that are more than 8 mm apart) per cluster.

4
5

1 **Table 4.** Relationship between intelligibility responses (Intelligible > Unintelligible

2 SWS) and the point at which participants reported recognizing that speech was

3 present

Run	Location	x	y	z	# Voxels	t	z	p_{FWE}
Run 1 & 2	L Inferior Frontal Gyrus	-58	14	22	189	4.81	4.05	.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	.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	.046
	R Superior Frontal Gyrus	14	52	8		4.81	4.05	
	R Anterior Cingulate Cortex	16	40	10		4.24	3.68	

4 Note. These results are presented at an uncorrected threshold of p < .001 peak level, FWE

5 corrected (p < .05) at cluster level. L = Left; R = Right. We report a maximum of 15 grey

6 matter local maxima (that are more than 8 mm apart) per cluster.

7
8
9

44