Comprehension of acoustically degraded speech in Alzheimer’s disease and primary progressive aphasia

Jessica Jiang¹, Jeremy CS Johnson¹, Maï-Carmen Requena-Komuro¹,², Elia Benhamou¹, Harri Sivasathiaseelan¹, Anthipa Chokesuwattanaskul¹,³, Annabel Nelson¹, Ross Nortley¹,⁴, Rimona S Weil¹, Anna Volkmer⁵, Charles R Marshall⁶, Doris–Eva Bamiou⁷, Jason D Warren¹*, Chris JD Hardy¹*

*joint senior authors

¹Dementia Research Centre, Department of Neurodegenerative Disease, UCL Queen Square Institute of Neurology, University College London, London, United Kingdom
²Kidney Cancer Program, UT Southwestern Medical Centre, Dallas, Texas, USA
³Division of Neurology, Department of Internal Medicine, King Chulalongkorn Memorial Hospital, Thai Red Cross Society, Bangkok, Thailand
⁴Wexham Park Hospital, Frimley Health NHS Foundation Trust, Berkshire, United Kingdom
⁵Division of Psychology and Language Sciences, University College London, London, United Kingdom
⁶Preventive Neurology Unit, Wolfson Institute of Population Health, Queen Mary University of London, London, United Kingdom
⁷UCL Ear Institute and UCL/UCLH Biomedical Research Centre, National Institute of Health Research, University College London, London, United Kingdom

Correspondence to: Dr Chris JD Hardy
Dementia Research Centre
UCL Institute of Neurology
1st Floor, 8-11 Queen Square
London WC1N 3AR, UK.
Email: chris.hardy@ucl.ac.uk
Tel: +44-203-448-3676

Word count: Main text 4,356, Abstract 354
Running title: Comprehension of degraded speech in dementia
Keywords: degraded speech, vocoded speech, auditory perception, primary progressive aphasia, frontotemporal dementia, Alzheimer’s disease

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
ABSTRACT

Successful communication in daily life frequently depends on accurate decoding of speech signals that are acoustically degraded by challenging listening conditions. This process presents the brain with a demanding computational task that is vulnerable to neurodegenerative pathologies. However, despite recent intense interest in the link between hearing impairment and dementia, daily hearing measures (such as degraded speech comprehension) in these diseases remain poorly defined. Here we addressed this issue in a cohort of 19 patients with typical Alzheimer’s disease (AD) and 31 patients representing canonical syndromes of primary progressive aphasia (PPA), in relation to 25 healthy age-matched controls. As a model paradigm for the acoustically degraded speech signals of daily life, we used noise-vocoding: synthetic division of the speech signal into a variable number of frequency channels constituted from amplitude-modulated white noise, such that fewer channels convey less spectrotemporal detail thereby reducing intelligibility. We investigated the impact of noise-vocoding on recognition of spoken three-digit numbers and used psychometric modelling to ascertain the threshold number of noise-vocoding channels required for 50% intelligibility by each participant. Associations of noise-vocoded speech intelligibility threshold with general demographic, clinical and neuropsychological characteristics and regional grey matter volume (defined by voxel-based morphometry of patients’ brain MR images) were also assessed. Compared with healthy older controls, all patient groups had a significantly higher mean noise-vocoded speech intelligibility threshold, particularly marked in logopenic variant and nonfluent-agrammatic variant PPA and significantly higher in AD than in semantic variant PPA (all p<0.05). Noise-vocoded intelligibility threshold discriminated dementia syndromes (in particular, Alzheimer’s disease) well from healthy controls. Further, this central hearing measure correlated with overall disease severity but not with measures of peripheral hearing or clear speech perception. Neuroanatomically, after correcting for multiple voxel-wise comparisons in pre-defined regions of interest, impaired noise-vocoded speech comprehension across dementia syndromes was significantly associated (p<0.05) with atrophy of left planum temporale, angular gyrus and anterior cingulate gyrus: a cortical network widely implicated in processing degraded speech signals. Taken together, our findings suggest that the comprehension of acoustically altered speech captures a central process relevant to daily hearing and communication in major dementia syndromes, with novel diagnostic and therapeutic implications.
INTRODUCTION

Successful communication in the world at large depends on our ability to understand spoken messages under non-ideal listening conditions. In our daily lives, we are required to interpret speech that is acoustically degraded by a wide variety of different ways – we regularly conduct conversations over background noise, adapt to suboptimal telephone and video connections and interpret unfamiliar accents. The processing of such degraded speech signals presents the brain with a challenging computational problem, whereby acoustic signals (or ‘auditory objects’) of interest must be disambiguated from interfering (and changing) noise.\(^1\)\(^-\)\(^3\) Because speech signals are critical for communication, decoding of degraded speech is generally the most functionally relevant index of hearing ability in daily life. This process, normally automatic and relatively effortless, is impaired in neurodegenerative disorders such as Alzheimer’s disease (AD) and the ‘language-led’ dementia syndromes of the primary progressive aphasia (PPA) spectrum.\(^4\)\(^-\)\(^8\)

Hearing impairment has recently been identified as a major risk factor for dementia and a driver of cognitive decline and disability.\(^4\)\(^,\)\(^9\)\(^,\)\(^10\) While most studies addressing this linkage have focused on peripheral hearing function measured using the detection of pure tones,\(^4\)\(^,\)\(^11\)\(^,\)\(^12\) mounting evidence suggests that measures of central hearing (auditory brain) function and in particular, the comprehension of degraded speech signals, may be more pertinent.\(^6\)\(^,\)\(^8\) Large cohort studies have identified impaired comprehension of degraded messages as a harbinger of dementia.\(^7\)\(^,\)\(^13\)\(^,\)\(^14\) More specifically, AD has been shown to impact speech-in-noise perception\(^15\) and identification of dichotic digits.\(^6\)\(^,\)\(^16\)\(^-\)\(^18\) This is likely to reflect, at least in part, a generic impairment of auditory scene analysis in AD, affecting the parsing of nonverbal as well as verbal information and linked to degeneration of the core temporo-parietal ‘default mode’ network targeted by AD pathology.\(^15\)\(^,\)\(^19\)\(^-\)\(^22\)

Further, both AD and PPA syndromes impair comprehension of non-native accents,\(^23\)\(^-\)\(^26\) sinewave speech\(^27\)\(^,\)\(^28\) and noise-interrupted speech,\(^29\) suggesting that neurodegenerative pathologies impair the processing of degraded speech signals more generally. However, the neural mechanisms responsible, the types of speech degradation that are implicated in everyday listening and the effects of different neurodegenerative pathologies have not yet been fully clarified. There are several grounds on which the processing of degraded speech may be especially vulnerable to neurodegenerative pathologies.\(^5\) Neuroanatomically, the processing of degraded speech signals engages distributed neural networks in peri-Sylvian, pre-frontal and posterior temporo-parietal cortices: these same brain networks are targeted preferentially in PPA, particularly the nonfluent/agrammatic variant (nfvPPA) and logopenic variant (lvPPA) syndromes.\(^5\)\(^,\)\(^27\)\(^,\)\(^30\)\(^,\)\(^31\) Computationally, the comprehension of degraded speech signals depends on precise yet dynamic
integration of information across neural circuitry \(^4,5,8,32,33\) and neurodegenerative pathologies are likely to blight these computations early and profoundly.

One widely used technique for altering speech signals experimentally is noise-vocoding, whereby a speech signal is divided digitally into discrete frequency bands (‘channels’), each filled with white noise and modulated by the amplitude envelope of the original signal.\(^34\) This procedure degrades the spectral content of the speech signal while preserving its overall longer range temporal structure. Using noise-vocoding, the level of intelligibility of the speech signal can be controlled: fewer channels is equivalent to less spectral detail available, leading to less intelligible speech. Among various alternative methods,\(^5\) noise-vocoding has certain attributes that make it attractive as a paradigm to study the effects of disease on the processing of degraded speech. Noise-vocoding has been widely studied and its behavioural and neuroanatomical correlates in the healthy brain are fairly well established.\(^34-40\) As an exemplar of acoustic degradation based on reduction of spectral information, it is broadly applicable to a variety of daily listening scenarios requiring decoding of ‘noisy’ speech signals (for example, a poor telephone or video-conferencing line). In contrast to speech-in-noise perception, comprehension of noise-vocoded speech depends intrinsically on auditory object (phonemic) decoding rather than selective attention. Further, noise-vocoding offers the substantial advantage of generating a quantifiable threshold for intelligibility of the degraded speech signal, based on the number of vocoding channels. This potentially allows for a more sensitive, graded and robust determination of deficit, enabling comparisons between diseases, tracking of disease evolution and assessing the impact of therapeutic interventions.

Noise-vocoding has been previously applied in a joint behavioural and magnetoencephalographic study of nfvPPA, to assess the brain mechanisms that mediate comprehension of degraded speech in the context of relatively focal cerebral atrophy.\(^41\) This work showed that patients with nfvPPA rely more on cross-modal cues to disambiguate vocoded speech signals, and have inflexible predictive decoding mechanisms, instantiated in left inferior frontal cortex. However, noise-vocoding has not been exploited as a tool to compare degraded speech perception in different neurodegenerative syndromes. More generally, the cognitive and neuroanatomical mechanisms that mediate the processing of degraded speech and their clinical resonance in this disease spectrum remain poorly defined.

Here, using noise-vocoding, we evaluated the comprehension of acoustically degraded spoken messages in cohorts of patients with typical AD and with all major syndromes of PPA, referenced to healthy older listeners. We assessed how the understanding of noise-vocoded speech was related to other demographic and disease characteristics. We further assessed the structural neuroanatomical associations of the noise-vocoded speech intelligibility threshold in AD and PPA,
using voxel-based morphometry on patients’ brain MR images. Based on available evidence with noise-vocoded and other degraded speech stimuli in these disease populations,\textsuperscript{5,30,31,41} we hypothesised that both AD and PPA patients would have elevated thresholds for comprehending vocoded speech compared with healthy controls, and that this deficit would be more severe in nfvPPA and lvPPA than in other neurodegenerative syndromes. We further hypothesised that elevated intelligibility threshold in the patient cohort would be correlated with regional grey matter atrophy in left posterior superior temporal, inferior parietal and inferior frontal cortices previously implicated in the processing of noise-vocoded speech in the healthy brain.\textsuperscript{34-40}

MATERIALS AND METHODS

Participants

Nineteen patients with typical amnestic AD, nine patients with lvPPA, 10 patients with nfvPPA and 12 patients with semantic variant primary progressive aphasia (svPPA) were recruited via a specialist cognitive clinic. All patients fulfilled consensus clinical diagnostic criteria with compatible brain MRI profiles and had clinically mild-to-moderate disease.\textsuperscript{42,43} No patients with pathogenic mutations were included. Twenty-five healthy older control participants with no history of neurological or psychiatric disorders were recruited from the Dementia Research Centre volunteer database. All participants had a comprehensive general neuropsychological assessment (Table 1). None had a history of otological disease, other than presbycusis; participants assessed in person at the research centre had pure tone audiometry, following a previously described procedure (details in Supplementary Material online).

Due to the Covid-19 pandemic, some data for this study were collected remotely (see Supplementary Materials). We have described the design and implementation of our remote neuropsychological assessment protocol elsewhere.\textsuperscript{44}

All participants gave informed consent to take part in the study. Ethical approval was granted by the UCL-NHNN Joint Research Ethics Committees, in accordance with Declaration of Helsinki guidelines.

Creation of experimental stimuli

Lists of 50 different three-digit numbers (of the form, ‘five hundred and eighty seven’; examples in Supplementary Material online) were recorded by two young adult female speakers in a Standard Southern British English accent with neutral prosody. They were recorded in Audacity (v 2.2.3), using a condenser microphone with a pop-shield in a sound-proof booth. Speech recordings were noise-vocoded using Matlab\textregistered{} (vR2019b) (https://uk.mathworks.com/) to generate acoustically altered stimuli with a prescribed level of degraded intelligibility (see Figure S1 for spectrograms).
Details concerning the synthesis of noise-vocoded stimuli are provided in Supplementary Material online. The vocoding intelligibility threshold for younger normal listeners is typically around three to four ‘channels’ \(^{34}\); in this experiment, we noise-vocoded the speech recordings with one to 24 channels, sampling at each integer number of channels within this range to ensure we would be able to accurately capture even markedly abnormal psychometric functions in the patient cohort.

The final stimulus list comprised 100 different spoken three-digit numbers: four unvocoded (clear speech) and 96 noise-vocoded with four stimuli for each number of channels, ranging from one to 24.

**Experimental procedure**

The stimuli were administered binaurally in a quiet room via Audio-Technica ATH-M50x headphones at a comfortable fixed listening level (at least 70 dB). Data for 30 participants were collected remotely via video link during the Covid-19 pandemic (see Table 1, further details in Supplementary Material online).

To familiarise them with the experimental procedure, participants were first asked to repeat five three-digit numbers (not included in the experimental session) that were spoken by the experimenter. Prior to presenting the experimental stimuli, participants were advised that the numbers they heard would vary in how difficult to understand they were, but that they should guess the number even if uncertain. Stimuli were presented in order of progressively decreasing channel number (intelligibility), from clear speech, then 24 to one vocoding channel. On each experimental trial, the task was to repeat the number (or as many of the three digits as the participant could identify). Participants were allowed to write down the numbers they heard rather than speaking them if preferred; in scoring, we accepted the intended target digit as correct, even if imperfectly articulated. Responses were recorded for offline analysis. During the experiment, no feedback about performance was given and no time limits were imposed.

**Analysis of clinical and behavioural data**

Data were analysed in Matlab® (vR2019b) and in R® (v4). For continuous demographic and neuropsychological data, participant groups were compared using ANOVA and Kruskal Wallis tests (dependent on normality of the data); group categorical data were compared using Fisher’s exact tests. Performance profiles in seven healthy control participants who performed the experiment both in person and subsequently remotely were very similar, justifying combining participants tested in person and remotely in the main analysis (see Supplementary Material online). An alpha of 0.05 was adopted as the threshold for statistical significance on all tests.
Identification of noise-vocoded spoken numbers was scored according to the number of digits correct for each three-digit number (e.g., if the target number was ‘587’ and the participant responded ‘585’, they would score two points on that trial). As three digits were presented on every trial, this scoring effectively yielded a total of 12 (4x3) data points for each vocoding channel number, for each participant. As the perceptual effect of noise-vocoding scales is exponential (so for example the increase in intelligibility for normal listeners is much greater between two and four channels than between 20 and 24 channels), we applied a logarithmic (base 2) transformation to the data. The resulting data were then modelled using a Weibull sigmoid, a widely used function for fitting logarithmically scaled data. Psychometric curves were created for individual participants and a mean curve was created for each diagnostic group using the Matlab psignifit package. For each function, we report the following parameters: the binaural noise-vocoded speech intelligibility threshold (the number of vocoding channels at which 50% identification of noise-vocoded numbers was achieved); the slope of the function at the threshold point; and lambda (the lapse rate, or number of incorrect responses at maximum performance asymptote).

As the data were not normally distributed, we used nonparametric Kruskal Wallis tests to analyse psychometric parameters. Where the omnibus test was significant, we conducted Dunn’s tests to conduct pairwise comparisons between participant groups. We assessed the relationship of noise-vocoded speech intelligibility threshold to forward digit span over the whole patient cohort, using Spearman’s correlation; here, digit span provides a metric of each patient’s overall ability to repeat (hear, hold in short term memory and articulate) natural spoken numbers. We further used Spearman’s correlation to assess the relationship of intelligibility threshold to general demographic (age, sex), clinical (symptom duration, Mini-Mental State Examination (MMSE) score), executive performance (WASI Matrices) and auditory perceptual (pure tone audiometry, phonemic pairs discrimination on the Psycholinguistic Assessment of Language Processing in Aphasia (PALPA)-3 subtest) measures, over the combined patient cohort.

Finally, receiver operating characteristic (ROC) curves were derived to assess the overall diagnostic utility of noise-vocoded speech comprehension in distinguishing each patient group from healthy controls. The binary classifier used was the 50% speech intelligibility threshold obtained from each psychometric function. The area under the ROC curve (AUC) was calculated for each syndromic group using parametric estimates in the pROC R package.

**Brain image acquisition and analysis**

Volumetric brain MR images were acquired for 25 patients in a 3 Tesla Siemens Prisma MRI scanner, using a 32-channel phased array head coil and following a T1-weighted sagittal 3D
magnetisation prepared rapid gradient echo (MPRAGE) sequence (TE = 2.9 ms, TI = 900 ms, TR = 2200 ms), with dimensions 256 mm x 256 mm x 208 mm and voxel size 1.1 mm x 1.1 mm x 1.1 mm.

For the VBM analysis, patients’ brain images were first pre-processed and normalised to MNI space using SPM12 software ([http://www.fil.ion.ucl.ac.uk/spm/software/spm12/](http://www.fil.ion.ucl.ac.uk/spm/software/spm12/)) and the DARTEL toolbox with default parameters running under Matlab R2014b. Images were smoothed using a 6-mm full-width at half-maximum Gaussian (FWHM) kernel. To control for individual differences in total (pre-morbid) brain size, total intracranial volume was calculated for each participant by summing white matter, grey matter and cerebrospinal fluid volumes post segmentation. An explicit brain mask was created using an automatic mask-creation strategy designed previously. A study-specific mean brain template image upon which to overlay statistical parametric maps was created by warping all patients’ native-space whole-brain images to the final DARTEL template and using the ImCalc function to generate an average of these images.

We assessed grey matter associations of noise-vocoded speech intelligibility threshold over the combined patient cohort. Voxel-wise grey matter intensity was modelled as a function of performance threshold in a multiple regression design, incorporating age, total intracranial volume and diagnostic group membership as covariates. Statistical parametric maps were assessed at peak-level significance threshold 

| RESULTS |

**General participant group characteristics**

Participant groups did not differ significantly in age, sex distribution, handedness or years of formal education (all p>0.05, Table 1). Patient groups did not differ in mean symptom duration (p=0.09) but did differ in MMSE score (H(3)=11.3, p=0.01; see Table 1), the AD group performing worse than the nfvPPA (z=-3.22, p=0.001) and svPPA (z=-2.10, p=0.04) groups. General neuropsychological profiles were in keeping with syndromic diagnosis for each patient group (Table 1). Pure tone audiometry (in the participant subcohort assessed in-person) revealed no substantial peripheral hearing deficits nor any significant differences between participant groups. Basic speech discrimination (assessed using the PALPA-3 phoneme discrimination subtest)
performance did not differ significantly from the healthy control group for any of the PPA syndromic groups.

**Experimental behavioural data**

Psychometric parameters for the participant groups are presented in Table 2 and noise-vocoded speech intelligibility thresholds for individual participants are plotted in Figure 1. Group mean psychometric functions are presented in Figure 2 and ROC curves for the patient groups versus the healthy control group in Figure 3. Exclusion of two upper bound outliers (>97.5 quantile) in parallel analyses left the results qualitatively unaltered. Results from the full dataset are accordingly reported in-text below; parallel analyses with outliers removed are reported in Supplementary Materials.

There was a significant main effect of diagnostic group on noise-vocoded speech intelligibility threshold (H(4) = 34.35, p<0.001). In post-hoc pairwise group comparisons versus healthy controls, mean intelligibility threshold was significantly elevated in all patient groups: in the lvPPA (z=3.87, p<0.001), nfvPPA (z=3.92, p<0.001), AD (z=5.01, p<0.001) and svPPA (z=2.20, p=0.03) groups. Comparing patient groups, intelligibility threshold was significantly elevated in the AD group compared with the svPPA group (z=2.04, p=0.04). There was no significant effect of diagnostic group on the slope of the psychometric function (p=0.247). There was however a significant main effect of diagnostic group on the lapse rate, lambda (H(4) = 16.75, p=0.002). In post-hoc pairwise group comparisons versus healthy controls, there was a significantly higher lapse rate (more errors made at maximum performance) in all patient groups: in the lvPPA (z=2.95, p=0.003), AD (z=2.61, p=0.009), nfvPPA (z=3.27, p=0.001), and svPPA (z=2.32, p=0.02) groups. There were no significant differences between patient groups for lapse rate.

Individual variability in psychometric parameters within participant groups was substantial (Figure 1, Table 2). Most pertinently, variation in noise-vocoded speech intelligibility threshold was wider in the AD group than in healthy controls and most marked in the lvPPA and nfvPPA groups.

Over the combined patient cohort, noise-vocoded speech intelligibility threshold was not significantly correlated with peripheral hearing function (r=-0.04, p=0.856), phonological discrimination in clear speech (PALPA-3 score; r=-0.25, p=0.185), age (r=0.24, p=0.100) or symptom duration (r=-0.10, p=0.510). Intelligibility threshold in the patient cohort was significantly correlated with WASI Matrices score (r=-0.49, p<0.001), MMSE score (r=-0.53, p<0.001) and forward digit span (r=-0.66, p<0.001). Lapse rate was also significantly correlated with forward digit span across the combined patient cohort (r=-0.34, p=0.017).
Analysis of ROC curves revealed that noise-vocoded speech intelligibility threshold discriminated all patient groups well from healthy controls. Based on AUC values (where a value of 1 would indicate an ideal classifier and values >0.8 a clinically robust discriminator), discrimination was ‘excellent’ for the AD group (AUC 0.95) and nfvPPA group (AUC 0.91), ‘good’ for the lvPPA group (AUC 0.88), and ‘fair’ for the svPPA group (AUC 0.77).

Neuroanatomical data
Statistical parametric maps of grey matter regions associated with speech intelligibility threshold are shown in Figure 4 and local maxima are summarised in Table 3.

Across the combined patient cohort, intelligibility threshold was significantly negatively associated with regional grey matter volume (i.e., associated with grey matter atrophy) in left planum temporale, left angular gyrus, and anterior cingulate gyrus (all \(p_{\text{FWE}} < 0.05\) after correction for multiple voxel-wise comparisons within the relevant pre-specified neuroanatomical region of interest).

DISCUSSION
Here we have shown that perception of acoustically degraded (noise-vocoded) speech is impaired in patients with AD and PPA syndromes relative to healthy older listeners, and further, stratifies syndromes: impairment was most severe in lvPPA and nfvPPA, and significantly more severe in AD than in svPPA. Intelligibility threshold for noise-vocoded speech did not correlate with measures of pure tone detection or phoneme discrimination in clear speech, suggesting that the deficit does not simply reflect a problem with peripheral hearing or elementary speech perception. Individual noise-vocoded speech intelligibility threshold varied widely within the AD, lvPPA and nfvPPA groups. Our findings suggest that elevation in noise-vocoded speech intelligibility threshold in these dementia syndromes captures a central auditory impairment potentially relevant to difficulties in diverse everyday listening situations requiring the decoding of acoustically altered speech signals.

Neuroanatomically, impaired noise-vocoded speech comprehension across dementia syndromes was underpinned by atrophy of left planum temporale, angular gyrus and anterior cingulate gyrus. This cortical network has been shown to be critical for processing speech signals under a range of noisy, daily listening conditions. Planum temporale is likely to play a fundamental role in the deconvolution of complex sound patterns and engagement of neural representations corresponding to phonemes and other auditory objects. Angular gyrus mediates the disambiguation of speech signals in challenging listening environments, working memory for speech signals and transcoding of auditory inputs for motor responses, including orienting and
Both regions are targeted in AD, lvPPA and nfvPPA and have been particularly implicated in the pathogenesis of impaired speech perception in these diseases. The anterior cingulate cortex operates in concert with these more posterior cortical hubs to decode spoken messages under challenging listening conditions, with a more general role in cognitive control and in allocating attentional resources to salient stimuli. Reduced activation of the anterior cingulate cortex during tracking of information in degraded speech signals has been demonstrated in nfvPPA and svPPA.

These neuroanatomical considerations suggest that the mechanisms of impaired noise-vocoded speech intelligibility are likely to differ between neurodegenerative syndromes, in keeping with the dissociable processes involved in phoneme recognition. Noise-vocoding fundamentally reduces the availability of acoustic cues that define phonemes as auditory objects: impaired recognition of these degraded auditory objects could in principle result from deficient encoding of acoustic features, damaged object-level representations (the auditory analogue of ‘apperceptive’ deficits in the visual domain) or impaired top-down, predictive disambiguation based on stored knowledge about speech signal characteristics. In AD and lvPPA, a core deficit of object-level representations has been demonstrated neuropsychologically and electrophysiologically using other procedures that alter acoustic detail in phonemes and nonverbal sounds; it is therefore plausible that an analogous apperceptive deficit may have impacted the recognition of noise-vocoded phonemes in the AD and lvPPA groups here. In nfvPPA, one previous study of noise-vocoded speech perception has foregrounded the role of inflexible top-down predictive decoding mechanisms instantiated in frontal cortex. However, this is a clinically, neuroanatomically and neuropathologically diverse syndrome, and involvement of posterior superior temporal cortex engaged in early auditory pattern analysis may constitute a ‘second hit’ to phoneme recognition. In svPPA, the elevated noise-vocoded intelligibility threshold is a priori more likely to reflect reduced activation of semantic mechanisms engaged in the predictive disambiguation of degraded speech signals; and indeed, comprehension of other kinds of acoustically degraded speech signals by patients with svPPA has previously been shown to be sensitive to semantic predictability and to engage anterior cingulate cortex.

Increasing intelligibility threshold was correlated with digit span over the combined patient cohort. This suggests that verbal working memory limitations may be integrally related to impaired processing of degraded speech, consistent with previous work highlighting the role of working memory in speech perception, particularly in older adults. As working memory demands did not vary across trials and number of vocoding channels, the principal driver of intelligibility threshold is likely to have been the level of acoustic alteration in the speech signal. On the other hand, all
patient groups showed an increased lapse rate (i.e., errors unrelated to the stimulus level\textsuperscript{45}) at higher vocoding channel numbers (i.e., for minimally noise-vocoded speech signals approaching clear speech). This echoes previous work demonstrating that active listening can be abnormal in lvPPA and nfvPPA even for clear speech and other sounds in quiet.\textsuperscript{65,74} As lapse rate was also correlated with digit span, reduced working memory may well have played a role here, potentially interacting with top-down mechanisms engaged in the predictive processing of speech.\textsuperscript{41} Indeed, frontal processes are likely to play a broader role in the disambiguation of degraded speech signals, including the allocation of attentional and executive resources\textsuperscript{75} and according with the observed correlation here between noise-vocoded speech intelligibility threshold and WASI Matrices score. Taken together, the present findings corroborate the profiles of deficit previously documented in AD and PPA syndromes for comprehension of sinewave speech and phonemic restoration in noise-interrupted speech.\textsuperscript{27,29}

Our findings further suggest that markers of noise-vocoded speech comprehension may have diagnostic and biomarker potential. The ROC analysis on the noise-vocoded intelligibility threshold measure (Figure 3) suggests that it would constitute an ‘excellent’ clinical test (corresponding to AUC > 0.9) for discriminating patients with AD and nfvPPA from healthy older individuals.\textsuperscript{55} Additionally, the noise-vocoded intelligibility threshold was correlated with overall disease severity (MMSE score) in the patient cohort. These findings build on a growing body of work suggesting that markers of ‘central’ hearing (auditory cognition) may sensitively signal the functional integrity of cortical regions that are vulnerable to AD and other neurodegenerative pathologies.\textsuperscript{5,8,14} The results of this study could further motivate the development of tailored strategies to help manage hearing difficulties experienced by people with dementia in various daily-life contexts and environments.

This study has limitations that suggest directions for further work. Our noise-vocoding paradigm (based on a step-wise linear progression through channel numbers) was not optimally efficient; an adaptive staircase procedure would reduce testing time and allow individual thresholds to be captured without administering uninformative trials at higher channel numbers. It would be relevant to assess to what extent patients’ comprehension of noise-vocoded speech can be modulated: pharmacologically (in particular, by acetylcholinesterase inhibitors\textsuperscript{28}) and/or by perceptual learning, as in healthy listeners.\textsuperscript{76-78} Using another kind of speech degradation (sinewave transformation), we have previously shown that pharmacological and perceptual learning effects may operate in AD and PPA syndromes.\textsuperscript{27,28} To establish how noise-vocoded speech perception and its modulatory factors relate to neural circuit integrity in AD and PPA, functional neuroimaging using techniques such as
fMRI and magnetoencephalography will be required to capture dynamic network connectivity engaged by these processes.

From a clinical perspective, this work should be taken forward in several ways. The group sizes here were relatively small: the noise-vocoding paradigm should be extended to larger patient cohorts, which (given the comparative rarity of PPA) will likely entail multi-centre collaboration. Besides corroborating the present group findings, assessment of larger cohorts would allow characterisation of the sources of the wide individual variation within diagnostic groups. There is also a need for prospective, longitudinal studies – both to assess how markers of degraded speech perception relate to disease course and to determine how early such markers may signal underlying neurodegenerative pathology. Auditory measures based on degraded speech comprehension would be well suited to future digital applications and potentially to large-scale screening of populations at risk of incident AD, as well as outcome measures in clinical trials of pharmacotherapies and non-pharmacological interventions.\textsuperscript{8,14} The key next step, however, will be to establish how well measures of degraded speech comprehension correlate with daily-life hearing and communication in AD and other neurodegenerative diseases – using both currently standardised symptom questionnaires and bespoke instruments developed to capture functional hearing disability in dementia. We have previously shown that pure tone audiometry alone is a poor predictor of everyday hearing\textsuperscript{79} while degraded speech performance may have better predictive value in patients with dementia.\textsuperscript{80} There would be considerable clinical value in a quantifiable index of degraded speech perception that could serve as a proxy and predictor of daily life hearing function and disability in major dementias: comprehension of noise-vocoded speech is a promising candidate.

The link between hearing impairment and dementia continues to be debated but presents a major opportunity for earlier diagnosis and intervention. Our findings suggest that the perception of degraded (noise-vocoded) speech captures quantifies central hearing functions beyond sound detection in dementia and stratifies major dementia syndromes. This central hearing index may constitute a proxy for the communication difficulties experienced by patients with AD and PPA under challenging listening conditions in daily life. We hope that this work will motivate further studies to define the diagnostic and therapeutic scope of central hearing measures based on degraded speech perception in these diseases.
ACKNOWLEDGEMENTS

We are grateful to all participants for their involvement. We thank Stuart Rosen for helpful advice on the application and analysis of the noise-vocoding paradigm.

FUNDING

The Dementia Research Centre is supported by Alzheimer’s Research UK, Brain Research Trust, and The Wolfson Foundation. The work was supported by the Alzheimer’s Society (grant AS-PG-16-007 to JDW), the Royal National Institute for Deaf People, Alzheimer’s Research UK and the National Institute for Health Research University College London Hospitals Biomedical Research Centre. This research was funded in part by the Wellcome Trust (grant no. 102129/B/13/Z) and UK Research and Innovation. For the purpose of Open Access, the authors have applied a Creative Commons Attribution (CC BY) public copyright licence to any Author Accepted Manuscript version arising from this submission. JJ is supported by a Frontotemporal Dementia Research Studentship in Memory of David Blechner (funded through The National Brain Appeal). JCSJ was supported by an Association of British Neurologists Clinical Research Training Fellowship. MCRK was supported by a Wellcome Trust PhD studentship (102129/B/13/Z). EB was supported by a Brain Research UK PhD Studentship. HS was funded by a Clinical Research Fellowship from the Leonard Wolfson Experimental Neurology Centre. AV is supported by an NIHR Advanced Fellowship (NIHR302240). CRM is supported by a grant from Bart’s Charity and the National Institute for Health Research. RSW is supported by a Wellcome Clinical Research Career Development Fellowship (205167/Z/16/Z). DEB is supported by the Royal National Institute for Deaf People. CJDH acknowledges funding from a RNID-Dunhill Medical Trust Pauline Ashley Fellowship (grant PA23_Hardy), a Wellcome Institutional Strategic Support Fund Award (204841/Z/16/Z) and the National Institute for Health Research.

COMPETING INTERESTS

The authors report no competing interests.
REFERENCES


82. Finger H, Goeke C, Diekamp D, Standvoß K, König P. LabVanced: a unified JavaScript framework for online studies. 2017:

# TABLES AND FIGURES

Table 1. General demographic, clinical and neuropsychological characteristics of all participant groups

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Controls</th>
<th>AD</th>
<th>lvPPA</th>
<th>nfvPPA</th>
<th>svPPA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographic and clinical</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. M:F</td>
<td>14:11</td>
<td>15:4</td>
<td>8:1</td>
<td>8:2</td>
<td>7:5</td>
</tr>
<tr>
<td>Age, years (y)</td>
<td>68.28 (6.62)</td>
<td>70.11 (8.43)</td>
<td>70.4 (6.46)</td>
<td>72.7 (3.65)</td>
<td>63.08 (8.38)</td>
</tr>
<tr>
<td>Handedness (R/L/A)</td>
<td>21/1/1</td>
<td>18/1/0</td>
<td>8/1/0</td>
<td>10/0/0</td>
<td>11/1/0</td>
</tr>
<tr>
<td>Education (y)</td>
<td>16.13 (2.72)</td>
<td>15.44 (3.79)</td>
<td>15.00 (3.12)</td>
<td>15.10 (2.64)</td>
<td>15.55 (2.07)</td>
</tr>
<tr>
<td>Symptom duration (y)</td>
<td>NA</td>
<td>5.94 (3.02)</td>
<td>6.62 (5.40)</td>
<td>3.30 (1.16)</td>
<td>5.42 (2.50)</td>
</tr>
<tr>
<td>Best ear average*</td>
<td>17.10 (8.72)</td>
<td>27.67 (10.92)</td>
<td>19.00 (10.70)</td>
<td>29.25 (3.30)</td>
<td>23.75 (8.07)</td>
</tr>
<tr>
<td>Tested in-person/remote</td>
<td>21/4</td>
<td>10/9</td>
<td>3/6</td>
<td>4/6</td>
<td>8/4</td>
</tr>
<tr>
<td>MMSE (/30)</td>
<td>29.75 (0.62)</td>
<td>20.43 (7.81)</td>
<td>22.67 (7.51)</td>
<td>26.50 (0.71)</td>
<td>22.88 (5.14)</td>
</tr>
<tr>
<td>T-MMSE (/27)</td>
<td>26.11 (1.76)</td>
<td>17.75 (4.45)</td>
<td>21.50 (4.51)</td>
<td>24.33 (2.25)</td>
<td>24.00 (1.41)</td>
</tr>
<tr>
<td>Percentage of participants</td>
<td>N/A</td>
<td>83.33%</td>
<td>83.33%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>taking donepezil and/or memantine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>General neuropsychology</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Executive function</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WASI Matrices (/32)</td>
<td>26.81 (2.74)</td>
<td>11.80 (8.76)</td>
<td>23.10 (5.44)</td>
<td>19.10 (9.70)</td>
<td>24.08 (6.73)</td>
</tr>
<tr>
<td>Letter fluency (total)</td>
<td>15.93 (5.35)</td>
<td>10.94 (5.86)</td>
<td>8.88 (3.98)</td>
<td>9.00 (9.20)</td>
<td>7.42 (6.07)</td>
</tr>
<tr>
<td>Category fluency (total)</td>
<td>24.07 (6.30)</td>
<td>11.41 (6.74)</td>
<td>11.10 (6.36)</td>
<td>15.43 (11.59)</td>
<td>6.67 (5.71)</td>
</tr>
<tr>
<td><strong>Working memory</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digit span forward (max)</td>
<td>6.56 (1.03)</td>
<td>5.79 (1.40)</td>
<td>4.56 (1.42)</td>
<td>5.60 (1.27)</td>
<td>6.67 (0.99)</td>
</tr>
<tr>
<td>Digit span reverse (max)</td>
<td>5.19 (1.17)</td>
<td>3.21 (1.40)</td>
<td>3.78 (1.30)</td>
<td>3.80 (1.93)</td>
<td>4.92 (1.56)</td>
</tr>
<tr>
<td><strong>Auditory input processing</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PALPA-3 (/36)</td>
<td>34.62 (1.66)</td>
<td>NA</td>
<td>31.20 (5.78)</td>
<td>31.40 (5.46)</td>
<td>33.67 (2.27)</td>
</tr>
<tr>
<td>Polyssyabic words (/45)</td>
<td>44.00 (1.55)</td>
<td>NA</td>
<td>41.30 (3.77)</td>
<td>38.89 (8.02)</td>
<td>40.67 (5.42)</td>
</tr>
<tr>
<td>Short sentences (/10)</td>
<td>9.46 (0.88)</td>
<td>NA</td>
<td>5.29 (1.60)</td>
<td>6.60 (2.80)</td>
<td>7.50 (2.11)</td>
</tr>
<tr>
<td><strong>Other language skills</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GNT (/30)</td>
<td>25.75 (2.46)</td>
<td>13.00 (7.22)</td>
<td>10.90 (6.94)</td>
<td>18.40 (8.82)</td>
<td>1.42 (4.32)</td>
</tr>
<tr>
<td>BPVS (/150)</td>
<td>147.88 (2.09)</td>
<td>135.32 (23.4)</td>
<td>146.00 (3.25)</td>
<td>127.50 (46.3)</td>
<td>73.50 (50.8)</td>
</tr>
<tr>
<td>PALPA-55 (/24)</td>
<td>23.46 (1.20)</td>
<td>NA</td>
<td>19.00 (4.47)</td>
<td>19.80 (5.12)</td>
<td>18.80 (6.09)</td>
</tr>
<tr>
<td><strong>Episodic memory</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMT Faces (Short) (/25)</td>
<td>23.75 (2.5)</td>
<td>16.11 (3.26)</td>
<td>21.40 (3.65)</td>
<td>22.83 (3.49)</td>
<td>19.20 (3.69)</td>
</tr>
<tr>
<td>RMT Faces (Long) (/50)</td>
<td>41.67 (3.70)</td>
<td>29.56 (5.57)</td>
<td>29.00 (6.93)</td>
<td>35.50 (4.95)</td>
<td>30.88 (3.36)</td>
</tr>
<tr>
<td><strong>Other skills</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDA calculation (/24)</td>
<td>14.81 (5.18)</td>
<td>5.06 (4.81)</td>
<td>6.14 (4.41)</td>
<td>7.00 (5.52)</td>
<td>10.18 (6.35)</td>
</tr>
<tr>
<td>VOSP Object Decision (/20)</td>
<td>18.94 (1.48)</td>
<td>14.10 (3.60)</td>
<td>17.00 (1.41)</td>
<td>16.44 (4.59)</td>
<td>15.50 (3.78)</td>
</tr>
</tbody>
</table>

Mean (standard deviation) values and raw scores are presented (maximum value possible in parentheses), unless otherwise indicated; significant differences from healthy controls (p<0.05) are in bold; significantly different to AD (p<0.05); significantly different to svPPA (p<0.05). *See Supplementary Material for details concerning the ‘best ear average’ measure. Participants assessed in-person did the MMSE while those assessed remotely did the T-MMSE. Similarly, the RMT
Faces (Long) was administered to participants in-person, versus the RMT Faces (Short), which was administered to participants remotely (full details of our remote neuropsychological test battery are given in 44. A, ambidextrous; AD, patient group with typical Alzheimer’s disease; BNT, Boston Naming Test; BPVS, British Picture Vocabulary Scale; Controls, healthy older control group; Digit span forward/reverse, maximum digit span recorded; F, female; GDA, Graded Difficulty Arithmetic; GNT, Graded Naming Test; L, left; lvPPA, patient group with logopenic variant primary progressive aphasia; M, male; MMSE, Mini-Mental State Examination; NA, not available/applicable; nfvPPA, patient group with nonfluent/agrammatic variant primary progressive aphasia; PALPA, Psycholinguistic Assessments of Language Processing in Aphasia; R, right; RMT, Recognition Memory Test; svPPA, patient group with semantic variant primary progressive aphasia; Synonyms concrete/abstract, single-word comprehension of single words; T-MMSE, tele-MMSE; VOSP, Visual Object and Space Perception battery; WASI, Wechsler Abbreviated Scale of Intelligence. a missing data for 1 participant; b missing data for 2 participants; c missing data for 3 participants; d missing data for 4 participants; e missing data for 6 participants; f missing data for 7 participants; g missing data for 8 participants; h missing data for 9 participants; i missing data for 10 participants; j missing data for 12 participants; k missing data for 13 participants; l missing data for 15 participants; m missing data for 21 participants.
### Table 2. Psychometric function parameters for comprehension of noise-vocoded speech in each participant group

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Controls</th>
<th>AD</th>
<th>lvPPA</th>
<th>nfvPPA</th>
<th>svPPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threshold</td>
<td>3.14 (0.27)</td>
<td>4.33 (1.25)*</td>
<td>5.04 (2.20)</td>
<td>4.68 (2.22)</td>
<td>3.55 (0.53)*</td>
</tr>
<tr>
<td>Slope</td>
<td>1.08 (0.90)</td>
<td>0.79 (0.27)</td>
<td>0.77 (0.36)</td>
<td>0.81 (0.27)</td>
<td>0.95 (0.46)</td>
</tr>
<tr>
<td>Lambda</td>
<td>0.00 (0.01)</td>
<td>0.02 (0.02)</td>
<td>0.03 (0.03)</td>
<td>0.02 (0.02)</td>
<td>0.02 (0.03)</td>
</tr>
</tbody>
</table>

Parameters are based on mean psychometric functions for each participant group (see text and Figure 2); mean (standard deviation) parameter values are shown. Threshold indicates 50% intelligibility of noise-vocoded spoken numbers, adjusted to take account of lambda value; slope indicates the slope of the psychometric function at this threshold point; lambda (lapse rate) indicates the number of incorrect responses at the maximum performance level. Significant differences (p<0.05) between patient groups and the healthy older control group are shown in bold; *, significantly lower in the svPPA group than the AD group (all p<0.05). AD, patient group with typical Alzheimer’s disease; Controls, healthy older control group; lvPPA, patient group with logopenic variant primary progressive aphasia; nfvPPA, patient group with nonfluent/agrammatic variant primary progressive aphasia; svPPA, patient group with semantic variant primary progressive aphasia.
Table 3. Neuroanatomical associations of noise-vocoded speech intelligibility threshold in the patient cohort

<table>
<thead>
<tr>
<th>Region</th>
<th>Cluster size (voxels)</th>
<th>Peak (mm)</th>
<th>T score</th>
<th>P FWE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>x</td>
<td>y</td>
<td>z</td>
</tr>
<tr>
<td>Left planum temporale</td>
<td>131</td>
<td>-48</td>
<td>-31</td>
<td>6</td>
</tr>
<tr>
<td>Left angular gyrus</td>
<td>36</td>
<td>-51</td>
<td>-61</td>
<td>16</td>
</tr>
<tr>
<td>Left cingulate gyrus</td>
<td>142</td>
<td>-1</td>
<td>38</td>
<td>-5</td>
</tr>
</tbody>
</table>

The table shows significant negative associations between regional grey matter volume and intelligibility threshold for noise-vocoded speech, based on the voxel-based morphometric analysis of brain MR images for the combined patient cohort. Coordinates of peaks (local maxima) are in MNI standard space. Local maxima shown were significant (p < 0.05) after family-wise error (FWE) correction for multiple voxel-wise comparisons within the pre-specified anatomical regions of interest (see text and Figure S2).
Figure 1. Plots of individual participant thresholds for comprehension of noise-vocoded speech within each diagnostic group. Speech intelligibility threshold values are based on individual psychometric curves for identification of noise-vocoded spoken numbers (see text for details). In this context, threshold corresponds to number of vocoding channels in the speech stimulus at which 50% intelligibility of spoken numbers was achieved, adjusted to take account of lambda value (the upper performance asymptote; see Table 2). The line within each box indicates the median, with the boxes indicating the interquartile interval. AD, patient group with typical Alzheimer’s disease; Control, healthy older control group; lvPPA, patient group with logopenic variant primary progressive aphasia; nfvPPA, patient group with nonfluent variant primary progressive aphasia; svPPA, patient group with semantic variant primary progressive aphasia.
Figure 2. Average psychometric curves for comprehension of noise-vocoded speech in each participant group. The y-axis here shows the percentage of digits identified correctly (from a total of 12 digits) at each noise-vocoding level; the x-axis shows the number of vocoding channels, plotted on a log scale. Mean psychometric functions were created for each diagnostic group (colour coded at lower right; see also text and Table 2); curves have been fitted through values (coloured dots) representing the mean score correct across individual participants in that group at each noise-vocoding level. AD, patient group with typical Alzheimer’s disease; Control, healthy older control group; lvPPA, patient group with logopenic variant primary progressive aphasia; nfvPPA, patient group with nonfluent variant primary progressive aphasia; svPPA, patient group with semantic variant primary progressive aphasia.
Figure 3. ROC curves for comprehension of noise-vocoded speech in patient groups versus healthy older controls. Receiver operating characteristic (ROC) curves for each syndromic group versus the healthy older control group are shown; the binary classifier used was the speech intelligibility threshold obtained in the psychometric functions (see Table 2 and Figure 2). An area under the curve (AUC) of 1 would correspond to an ideal classifier. AUC values obtained were as follows: Alzheimer’s disease, AUC = 95%; logopenic variant primary progressive aphasia (PPA), AUC = 88%; nonfluent/agrammatic variant PPA, AUC = 91%; semantic variant PPA, AUC = 77%. 
Figure 4. Statistical parametric maps of regional grey matter atrophy associated with elevated noise-vocoded speech intelligibility threshold in the combined patient cohort. Maps are rendered on sagittal sections of the group mean T1-weighted MR image in MRI space, masked using the pre-specified neuroanatomical region of interests (as used in the small volume corrections) and thresholded at $p < 0.001$ uncorrected for multiple voxel-wise comparisons over the whole brain for display purposes (areas shown were significant at $p < 0.05_{\text{FWE}}$ for multiple comparisons within regions of interest). The colour bar (right) codes voxel-wise t-values. All sections are through the left cerebral hemisphere; the plane of each section is indicated using the corresponding MNI coordinate (mm).