

# Voice processing in dementia: a neuropsychological and neuroanatomical analysis

Julia C. Hailstone,<sup>1</sup> Gerard R. Ridgway,<sup>1,2</sup> Jonathan W. Bartlett,<sup>1,3</sup> Johanna C. Goll,<sup>1</sup>  
Aisling H. Buckley,<sup>1</sup> Sebastian J. Crutch<sup>1</sup> and Jason D. Warren<sup>1</sup>

1 Dementia Research Centre, University College London Institute of Neurology, Queen Square, London WC1N 3BG, UK

2 Wellcome Trust Centre for Neuroimaging, University College London Institute of Neurology, Queen Square, London WC1N 3BG, UK

3 Department of Medical Statistics, London School of Hygiene & Tropical Medicine, London, UK

Correspondence to: Dr J.D. Warren,  
Dementia Research Centre,  
Institute of Neurology,  
University College London,  
Queen Square,  
London WC1N 3BG, UK  
E-mail: jwarren@drc.ion.ucl.ac.uk

Voice processing in neurodegenerative disease is poorly understood. Here we undertook a systematic investigation of voice processing in a cohort of patients with clinical diagnoses representing two canonical dementia syndromes: temporal variant frontotemporal lobar degeneration ( $n = 14$ ) and Alzheimer's disease ( $n = 22$ ). Patient performance was compared with a healthy matched control group ( $n = 35$ ). All subjects had a comprehensive neuropsychological assessment including measures of voice perception (vocal size, gender, speaker discrimination) and voice recognition (familiarity, identification, naming and cross-modal matching) and equivalent measures of face and name processing. Neuroanatomical associations of voice processing performance were assessed using voxel-based morphometry. Both disease groups showed deficits on all aspects of voice recognition and impairment was more severe in the temporal variant frontotemporal lobar degeneration group than the Alzheimer's disease group. Face and name recognition were also impaired in both disease groups and name recognition was significantly more impaired than other modalities in the temporal variant frontotemporal lobar degeneration group. The Alzheimer's disease group showed additional deficits of vocal gender perception and voice discrimination. The neuroanatomical analysis across both disease groups revealed common grey matter associations of familiarity, identification and cross-modal recognition in all modalities in the right temporal pole and anterior fusiform gyrus; while in the Alzheimer's disease group, voice discrimination was associated with grey matter in the right inferior parietal lobe. The findings suggest that impairments of voice recognition are significant in both these canonical dementia syndromes but particularly severe in temporal variant frontotemporal lobar degeneration, whereas impairments of voice perception may show relative specificity for Alzheimer's disease. The right anterior temporal lobe is likely to have a critical role in the recognition of voices and other modalities of person knowledge.

**Keywords:** voice; face; phonagnosia; frontotemporal dementia; Alzheimer's disease

**Abbreviations:** FTD = frontotemporal dementia; FTLD = frontotemporal lobar degeneration

## Introduction

Disorders of face processing are well recognized and widely studied, but much less attention has been paid to disorders of voice processing. From a neurobiological perspective this is somewhat surprising, since human voices can be considered ‘auditory faces’ in several important respects (Schweinberger *et al.*, 1997a; Belin *et al.*, 2004). Ethologically, voices are the second major channel via which we identify other people, while perceptually, individual voices (like faces) present the brain with a formidable problem of precise differentiation and identification within a class of complex sensory objects. The relative neglect of voice processing in the clinical literature is likely to reflect a lack of theoretical frameworks for understanding voice analysis, the comparative difficulty of working with vocal stimuli and the scarcity of symptomatic deficits of voice processing (phonagnosia), in turn partly attributable to the primacy of the face as a source of person knowledge in daily life. However, this situation has been transformed recently with the development of cognitive models of voice processing (Belin *et al.*, 2000, 2004; von Kriegstein *et al.*, 2005, 2006; Campanella and Belin, 2007; Goll *et al.*, 2010), the availability of sophisticated stimulus synthesis and delivery systems, functional imaging studies in the healthy brain and the detailed characterization of phonagnosias. Cognitive models of voice processing have been influenced by face processing models (Bruce and Young, 1986; Burton *et al.*, 1990; Ellis *et al.*, 1997). In the voice processing model proposed by Belin *et al.* (2004), voice identification occurs via serially and hierarchically organized processing stages: perceptual representations or templates derived from modality-specific (voice and face) recognition units feed into amodal ‘person identity nodes’, evoking a sense of familiarity that gates access to biographical information (including personal names) and leads to recognition of the speaker. These putative processing stages map onto an anatomical hierarchy in which perceptual analysis of voices occurs in temporoparietal cortices with associative processing of voices and other modalities of person knowledge in more anterior temporal lobe areas, delineated using functional brain imaging (Imaizumi *et al.*, 1997; Nakamura *et al.*, 2001; Belin *et al.*, 2002; von Kriegstein *et al.*, 2006; Warren *et al.*, 2006; Bishop and Miller, 2009). However, the cognitive and neural architecture of voice processing and its relations to other modalities of person knowledge continue to be defined. Phonagnosia has been described as a developmental disorder (Garrido *et al.*, 2009) and, more commonly, in association with focal damage involving the right or left temporal lobe or the right parietal lobe (Van Lancker and Canter, 1982; Van Lancker and Kreiman, 1987; Van Lancker *et al.*, 1988, 1989; Ellis *et al.*, 1989; Hanley *et al.*, 1989; Neuner and Schweinberger, 2000; Belin *et al.*, 2004; Lang *et al.*, 2009).

There are both clinical and neurobiological grounds for a systematic analysis of voice processing in the degenerative dementias. Clinically, voice processing impairments are likely to be under-recognized in these diseases yet may constitute a significant and disabling symptom (Hailstone *et al.*, 2010), especially in situations where additional cues to speaker identity are reduced or unavailable. Neurobiologically, the common dementias collectively affect

distributed areas in the temporal, parietal and frontal lobes that have been implicated in voice processing in functional imaging studies of healthy subjects (Seeley *et al.*, 2009). In particular, the brunt of tissue damage in Alzheimer’s disease and several diseases in the frontotemporal lobar degeneration (FTLD) spectrum initially falls on the temporal lobes, which are likely to contain mechanisms integral for voice analysis and recognition (Belin *et al.*, 2004). The syndrome of progressive prosopagnosia is well recognized in association with right temporal lobe atrophy (Evans *et al.*, 1995; Joubert *et al.*, 2003, 2004; Josephs *et al.*, 2008; Chan *et al.*, 2009) and is dominated by deficits of non-verbal knowledge, in particular knowledge of familiar people (Hanley *et al.*, 1989; Gentileschi *et al.*, 1999, 2001; Snowden *et al.*, 2004; Thompson *et al.*, 2004; Gainotti, 2007, 2008). While voice recognition commonly becomes affected with evolution of the progressive prosopagnosia syndrome (Gentileschi *et al.*, 2001; Gainotti *et al.*, 2003, 2008), selective phonagnosia has seldom been reported. However, progressive associative phonagnosia with relatively-preserved voice discrimination, face and proper name recognition has recently been described in association with frontotemporal atrophy involving the right anterior temporal lobe (Hailstone *et al.*, 2010). Neurodegenerative diseases offer a perspective on voice processing that is complementary both to studies in normal subjects and in patients with focal brain lesions: the breakdown of voice processing in dementia would potentially allow identification of critical nodes in a functional and anatomical cerebral network, and inform cognitive models of voice processing.

In this study, we investigated neuropsychological and neuroanatomical signatures of voice processing in two canonical dementias, FTLD and Alzheimer’s disease. In targeting these disease groups we recognized that, whereas Alzheimer’s disease typically presents with a relatively uniform clinical and anatomical profile, FTLD is clinically and anatomically heterogeneous. In particular, patients with FTLD who have predominant temporal lobe atrophy (i.e. those predicted *a priori* to develop voice processing deficits) have heterogeneous clinical presentations, including both semantic dementia (progressive semantic aphasia or progressive prosopagnosia) and progressive behavioural decline [(behavioural variant frontotemporal dementia (FTD)]. For the purposes of the present ‘lesion-led’ study, we selected patients with FTLD based on the presence of predominant temporal lobe atrophy: we term this non-canonical, anatomically defined subgroup ‘temporal lobe variant FTLD’ (Brambati *et al.*, 2009). In line with current cognitive models and previous neuropsychological evidence concerning the organization of voice processing (Van Lancker *et al.*, 1988; Ellis *et al.*, 1997; Schweinberger *et al.*, 1997b; Neuner and Schweinberger, 2000; Belin *et al.*, 2004; Lucchelli and Spinnler, 2008; Garrido *et al.*, 2009; Hanley and Damjanovic, 2009), we designed a series of neuropsychological experiments incorporating subtests to assess early perceptual encoding, discrimination and recognition of voices in the target clinical groups. Processing of voices was assessed in relation to processing of faces and names, in order to assess the modality- and material-specificity of any voice processing deficit. Neuroanatomical correlates of voice processing performance were assessed using voxel-based morphometry (Ashburner and Friston, 2000). We hypothesized distinct

profiles of phonagnosia in temporal lobe variant FTL and Alzheimer's disease, with more severe associative impairment in temporal lobe variant FTL and relatively more prominent apperceptive impairment in Alzheimer's disease. We further hypothesized, based on anatomical evidence in the healthy brain, that semantic deficits in processing voices (in common with other kinds of person knowledge) would be associated with atrophy of anterior temporal lobe regions, and voice apperceptive deficits would be associated with atrophy of more posterior temporo-parietal regions.

## Materials and methods

### Subject details

Fourteen consecutive patients with temporal lobe variant FTL, 22 patients with Alzheimer's disease and 35 healthy older control subjects participated. All patients were recruited from the tertiary Cognitive Disorders Clinic at the National Hospital for Neurology and Neurosurgery. Patients with temporal lobe variant FTL were selected based on the presence of selective, bilateral anterior temporal lobe

atrophy on MRI (atrophy of one or both temporal lobes disproportionate to any accompanying atrophy of other cerebral regions, as assessed visually by an experienced, independent neuroradiologist); the distribution of temporal lobe atrophy was asymmetric in 13/14 cases. Clinically, most (13/14) patients with temporal lobe variant FTL had a syndrome of semantic dementia according to the consensus criteria of Neary *et al.* (1998); within this semantic dementia subgroup, 10 patients presented with progressive semantic aphasia and three presented with progressive prosopagnosia. One patient within the FTL group presented with behavioural variant FTD. All patients with Alzheimer's disease had a clinical syndrome of typical Alzheimer's disease led by memory decline and fulfilled modified National Institute of Neurological and Communicative Disorders and Stroke/Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria for probable Alzheimer's disease (Dubois *et al.*, 2007). Twenty out of 22 patients with a clinical diagnosis of Alzheimer's disease had brain MRI: 16 patients had disproportionate symmetrical hippocampal atrophy and four had generalized cerebral atrophy. Demographic characteristics of the study groups are summarized in Table 1 and further details are provided in the online Supplementary material. Significance of differences between groups was assessed using z-tests with bootstrap (2000 replicates) standard errors. Subject groups did not differ significantly in age, gender distribution or years of education (all

**Table 1** Summary of subject characteristics

	Temporal lobe variant FTL <i>n</i> = 14		Alzheimer's disease <i>n</i> = 22		Healthy controls <i>n</i> = 35	
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
Demographic characteristics						
Males: females	8:6		10:12		13:22	
Right: left-handed	11:3		19:3		31:4	
Age (years)	64.2 (6.3)	54–76	66.5 (7.7)	49–79	63.9 (5.7)	54–79
Years of education	13.9 (4.8)	10–25	13.5 (3.6)	9–20	15.2 (3.3)	11–25
Clinical characteristics						
Clinical syndrome at presentation	Semantic dementia ( <i>n</i> = 13) <sup>a</sup> Behavioural variant FTD ( <i>n</i> = 1)		Amnesic Alzheimer's disease		NA	
Symptom duration (years)	5.4 (1.7)	3–8	5.7 (2.4)	2–11	NA	
Mini-Mental State Examination (/30)	21.1 (7.2)**	6–29	21.3 (4.2)**	14–28	29.4 (0.6) <sup>b</sup>	28–30
Medication	4 <sup>c</sup>		18 <sup>d</sup>		NA	
Cardinal symptoms <sup>e</sup>						
Voice recognition	<i>n</i> = 9 (64%)		<i>n</i> = 11 (50%)		NA	
Face recognition	<i>n</i> = 7 (50%)		<i>n</i> = 8 (36%)		NA	
Voice familiarity	<i>n</i> = 3 (21%)		<i>n</i> = 3 (14%)		NA	
Face familiarity	<i>n</i> = 5 (36%)		<i>n</i> = 2 (9%)		NA	
Media exposure <sup>e</sup>						
TV watching (hours per week)	15.1 (9.2)	0–32	15.9 (10.0)	0–35	14.4 (10.5)	0–63
Radio listening (hours per week)	2.4 (4.2)**‡	0–13	11.4 (13.3)	0–42	13.8 (12.0)	0.5–55
News exposure (times per week)	8.1 (4.3) **†	1–20	13.0 (8.0)	1–30	13.9 (6.9)	4–35

Assessed using questionnaire (Supplementary material).

a Ten cases with progressive semantic aphasia, three cases with progressive prosopagnosia.

b Twenty-three controls performed a Mini-Mental State Examination.

c Two patients taking a serotonin reuptake inhibitor, one taking anti-Parkinson's medication, one taking lithium.

d Sixteen patients taking a cholinesterase inhibitor, two taking memantine.

e Assessed using questionnaire (Supplementary material).

NA Not applicable to controls.

\*Significantly worse than controls ( $P < 0.05$ ); \*\*significantly worse than controls ( $P < 0.001$ ); †significantly different from other patient group ( $P < 0.05$ ); ‡significantly different from other patient group ( $P < 0.01$ ).

$P > 0.05$ ); the temporal lobe variant FTL and Alzheimer's disease groups did not differ significantly on two general measures of clinical severity (symptom duration and Mini-Mental State Examination score).

The study was approved by the local institutional research ethics committee and all subjects gave informed consent in accord with the principles of the Declaration of Helsinki.

### General neuropsychological assessment

All patients and 19 healthy control subjects had a comprehensive general neuropsychological assessment; the remaining 16 control subjects performed a reduced set of tests. The tests administered are listed in Table 2. Fisher's exact test was used to assess group differences in gender, for all other variables differences in means were assessed using z-tests with bootstrap (2000 replicates) standard errors.

### Peripheral hearing assessment

Most subjects had no clinical history of hearing loss. To assess any effects of hearing loss on performance in the experimental tasks across the experimental groups, all subjects underwent pure tone audiometry

on frequencies between 0.5 and 4 kHz. The audiometry procedure and analysis are described in the online Supplementary material.

## Experimental behavioural investigations: perceptual analysis of voice attributes

Voices are complex auditory objects and voice perception depends on encoding spectro-temporal attributes that carry information about vocal identity such as gender and size (Belin *et al.*, 2002, 2004; Griffiths and Warren, 2004; Warren *et al.*, 2006). Here we created novel subtests to assess encoding of the voice attributes of gender and vocal tract length (an index of vocal size).

### Vocal gender

Vocal gender is determined by lower level perceptual properties including pitch and vocal tract length; here we were interested in subjects' ability to assign gender to vocal samples based on all available auditory

**Table 2 Results of general neuropsychological assessment**

Test (max score)	Temporal lobe variant FTL $n = 14$		Alzheimer's disease $n = 22$		Healthy controls $n = 35$	
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
IQ						
WASI Verbal IQ	67.6 (21.7)**‡	40–111	96.9 (17.2)**	67–121	120.8 (9.2)	100–141
WASI Performance IQ	99.9 (19.1)*	68–133	86.0 (16.3)**†	62–110	116.8 (11.9)	96–142
Reading IQ <sup>a</sup>	88.7 (23.9)**†	45–122	106.4 (15.7)*	68–128	118.9 (7.4)	96–129
Semantic tests						
BPVS (/150)	73.8 (49.7)**‡	5–148	141.4 (11.9)*	106–150	148.1 (1.5)	144–150
Concrete synonyms (/25)	14.2 (5.3)**†b	7–24	20.9 (2.7)**b	13–24	24.3 (1.3)	19–25
Abstract synonyms (/25)	15.4 (5.8) **†b	8–24	20.9 (3.5)**b	14–25	24.3 (1.2)	20–25
Landmark name (/15)	2.6 (3.7)**†c	0–12	6.1 (4.0)**c	0–15	13.5 (1.3) <sup>d</sup>	11–15
Landmark identify (/15)	4.6 (4.7)**†c	0–12	8.0 (4.1)**c	0–15	13.7 (1.2) <sup>d</sup>	11–15
Other non-semantic skills						
GNT (/30)	2.2 (6.1)**‡	0–23	11.6 (7.9)**	0–26	26.0 (2.4)	19–30
ODT (/20)	16.5 (5.0)	8–29	15.8 (2.8)**	9–19	18.5 (1.2)	16–20
Forwards DS (/12)	7.3 (2.7)	4–12	7.1 (2.3)*	4–11	8.7 (2.0)	4–12
Reverse DS (/12)	6.1 (3.3)	0–10	4.9 (2.7)*	0–10	7.4 (2.6)	2–12
GDA (/24)	8.9 (6.9)**	0–20	5.5 (4.5)**	0–14	15.4 (4.8) <sup>e</sup>	6–23
Episodic memory						
RMT words (/50)	35.4 (7.0)**h	24–47	30.1 (7.3)**†	19–47	47.3 (1.8) <sup>e</sup>	43–49
RMT faces (/50)	28.9 (4.1)**‡h	24–40	35.0 (5.6)**	25–45	42.2 (4.7) <sup>e</sup>	35–49
Executive function						
Stroop word reading: scaled score	5.2 (4.0) <sup>f</sup>	1–14	5.8 (4.6)**	1–13	10.7 (2.7) <sup>e</sup>	3–14
Stroop inhibition: scaled score	6.3 (4.6)**g	1–13	3.6 (3.2)**	1–11	11.5 (2.0) <sup>e</sup>	7–14

a Reading IQ measured on the National Adult Reading Test (NART) (Nelson, 1982) unless the subject scored  $\leq 15/50$  on this test, in which case the Schonell Graded Word Reading Test IQ was used (Schonell and Goodacre, 1971).

b Two temporal lobe variant FTL, one subject with Alzheimer's disease did not perform synonyms tests.

c Three subjects with temporal lobe variant FTL, two subjects with Alzheimer's disease did not perform the London landmarks test.

d 34/35 controls tested on these tasks.

e 19/35 controls tested on these tasks.

f One temporal lobe variant FTL subject unable to read the words and a scaled score of 1 was used.

g  $n = 12$  (two temporal lobe variant FTL subjects unable to name colours).

h One temporal lobe variant FTL subject did not perform recognition memory tasks.

\*Significantly worse than controls ( $P < 0.01$ ); \*\*significantly worse than controls ( $P < 0.001$ ); †significantly worse than other patient group ( $P < 0.05$ ); ‡significantly worse than other patient group ( $P < 0.001$ ).

BPVS = British Picture Vocabulary Scale (McCarthy and Warrington, 1992); Concrete and Abstract Synonyms Test (Warrington *et al.*, 1998); DS = Wechsler Memory Scale - Revised (WMS-R) digit span (Wechsler, 1987); GDA = Graded Difficulty Arithmetic; GNT = Graded Naming Test (Warrington, 1997); Landmark name/identify = London landmark naming and identification test (Whiteley and Warrington, 1978); ODT = Object Decision Task (Warrington and James, 1991); RMT = Recognition Memory Tests (Warrington, 1984); Stroop, Delis-Kaplan Executive Function System Stroop test (Delis *et al.* 2001); WASI = Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999).

cues. Vocal samples (each 5 s in duration) were derived from publicly available sources. Twenty-four trials (12 male) were presented; the task was to decide if the voice was male or female.

### Vocal size (vocal tract length)

Like voice gender, vocal size is determined by lower-level properties: in this subtest we were interested in subjects' ability to make more fine-grained categorical perceptual judgements of vocal size by manipulating vocal tract length information in isolation, as previously described in normal subjects (Ives *et al.*, 2005). Stimuli were based on 10 consonant–vowel syllables recorded by a single male speaker and digitally resynthesized using a previously described algorithm (Kawahara and Irino, 2004) that allows apparent vocal tract length to be varied independently of glottal pulse rate (voice pitch). Each syllable was presented at two extreme vocal tract length values, one corresponding to a speaker height of 2 m (equivalent to a very tall male, 'big') and the other to a height of 0.5 m (equivalent to a child, 'small'), and randomly assigned one of four pitches within the normal human male vocal range (116, 120, 172, 190 Hz), which was varied independently of vocal tract length. Examples of stimuli are available from the authors. Twenty trials (10 'big', 10 'small') were presented; on each trial, subjects heard a sequence of four repetitions of the same stimulus, and were asked to judge if the sounds were made by a big person or a small person.

## Experimental behavioural investigations: voice discrimination

Distinguishing between voices of comparable pitch and vocal tract length requires comparison between object-level representations of each voice: this processing stage beyond early perceptual analysis can be regarded as an auditory analogue of apperceptive processing in the visual domain (Goll *et al.*, 2010). We created a voice discrimination task in which subjects were required to detect a change in speaker within a spoken phrase. The verbal content of the phrase was a highly over-learned series ('Monday, Tuesday, Wednesday, Thursday'). All speakers were female, aged 21–31 years, with a standard Southern English accent. Two versions of the test stimuli were recorded to create two levels of speaker discrimination task difficulty: in the 'easy' version of the task, voice pitch ( $f_0$ ) was not fixed while in the 'difficult' version of the task, inter-speaker variations in vocal pitch were controlled by setting  $f_0$  of recorded stimuli at 220 Hz using Goldwave® software. Recorded single words were concatenated with fixed inter-word gap (0.1 s) to equate overall speech rate. If the sequence contained a speaker change, this change always occurred at the midpoint of the phrase, to maximize available vocal information for each speaker. Examples of stimuli are available from the authors. In the 'easy' discrimination test 28 trials (14 speaker fixed, 14 speaker change) were presented; in the 'difficult' speaker discrimination test, 12 trials (six speaker fixed, six speaker change) were presented. The task was to decide whether the spoken phrase contained a change in speaker.

Patient performance on these vocal tasks was compared with performance on a standard test of perceptual processing of face identity, the Benton Facial Recognition Test (Benton *et al.*, 1989): this test depends on successful perceptual encoding of the configuration of a face, and requires the subject to match a photograph of target face to one (or three) of six other photographs of the target with distractor faces under different viewing conditions. The short form of the test was administered. Scores ( $/56$ ) were normalized for age and education.

## Experimental behavioural investigations: voice recognition

In this test we aimed to assess different aspects of semantic processing of famous voices (voice recognition) comprising familiarity, naming, identification from biographical information and cross-modal matching, in comparison to recognition of faces and names for the same famous individuals. Sixty voice samples and face photographs were initially obtained from publicly available sources on the internet and chosen so as to minimize other potential semantic cues to recognition. A set of 24 voice samples of well-known public figures was selected from this larger set, based on a pilot analysis in a separate group of healthy older British controls (Hailstone *et al.*, 2010); the best-recognized samples were included in the final stimulus set used in all semantic subtests. These 24 famous individuals (Supplementary Table 1) comprised 10 politicians, five actors, seven other media personalities from television and radio, and two members of the British royal family. Examples of stimuli are available from the authors. Face photographs and names of the same set of 24 famous individuals were used for the face and name processing tasks, respectively.

### Familiarity of voices, faces and names

In this subtest the set of 24 famous voices was supplemented by 24 unfamiliar voices and faces [as classified by >75% of healthy controls in the pilot study: (Hailstone *et al.*, 2010)] matched by gender to the famous set and approximately matched for age and accent. The written names of the same 24 famous individuals were supplemented with 24 fabricated personal name foils. For each modality, 48 trials (24 famous, 24 unfamiliar) were presented; each stimulus was presented once, and the task was to decide if the stimulus was familiar in a forced choice ('yes – no') protocol. As voice familiarity was the primary focus here, voices were presented first (in order to minimize priming effects in the voice modality), followed by faces and names.

### Identification of voices, faces and names

In this subtest, the task was to identify the person as precisely as possible; for voices and faces, if the subject was not able to name the person, they were asked to provide other biographical details (e.g. an event closely associated with the person, occupational information), in line with the criteria used by Snowden *et al.* (2004). For names, on each trial the subject was required to provide identifying information about the person. For voices, national or regional origin was not accepted as evidence of person recognition, since this could be based on accent cues alone. As voice identification was the focus here, voices were presented first (in order to minimize priming effects in the voice modality), followed by faces and names.

### Cross-modal recognition of voices and faces

We employed a cross-modal matching task in order to allow patients to demonstrate recognition of voices and faces using an alternative procedure that did not rely on word retrieval. For both face and voice targets, three stimulus arrays were selected using individuals from the set of 24 individuals; each individual was represented in one of the arrays. The first array contained the six females from the complete set, a second array contained the nine male politicians, and the third contained the nine male media figures (as career is likely to be an important organizational principle in the domain of person knowledge; Crutch and Warrington, 2004). The set of 24 faces was presented first, and the task was to match the face to one of the

names in the array. The set of 24 famous voices was presented with the same arrays but with simultaneous presentations of faces and names in each array; the task was to match the voice to one of the face-name pairs.

## Experimental behavioural investigations: general procedure

The experimental tests were administered to subjects over several sessions. Auditory stimuli were presented from digital wavefiles on a laptop computer via headphones at a comfortable listening level in a quiet room. Visual stimuli were presented as high-quality greyscale photographs. Verbal stimuli were simultaneously presented as written words and spoken by the examiner (control subjects were presented with written words only). All stimuli were presented in fixed randomized order. Subject responses were collected for off-line analysis. Before beginning each test, several practice trials were administered to ensure the subject understood the task. No feedback about performance was given during the test. Voice stimuli were presented once only for all tasks, with the exception of the famous voice identification task, where subjects were permitted a further presentation if requested. No time limit was imposed.

## Experimental behavioural investigations: analyses

Experimental behavioural data were analysed in STATA R9.2 (Stata Corporation). For the perceptual tests, differences in mean scores between groups were assessed using z-tests and 95% Wald-type confidence intervals, with standard errors calculated using bootstrapping (2000 replicates).

For the semantic subtests, the effect of stimulus presentation modality was assessed using a bootstrapped linear regression model with 2000 replicates, which allowed for the repeated measures from subjects. A global Wald test of interaction was carried out to test the hypothesis that group differences in scores varied between modalities, and modality-associated differences in performance between the temporal variant FTLD and Alzheimer's disease groups were assessed in pair-wise comparisons between modalities. Using this model, differences between the two patient groups were adjusted for modality performance differences exhibited by healthy controls (anticipated from previous work; Hailstone *et al.*, 2010).

Within each patient group, correlation coefficients between experimental tests were estimated with 95% bias-corrected bootstrap confidence intervals (2000 replicates). Correlations were estimated between perceptual discrimination subtest scores; between semantic subtest scores within and between modalities; and between perceptual and semantic performance. Associations with disease severity measures were also assessed, using linear regression models with 95% Wald-type bootstrap confidence intervals with 2000 replicates. As severity measures, symptom (clinical disease) duration was used for both disease groups; in addition, Mini-Mental State Examination score was used for the Alzheimer's disease group and the British Picture Vocabulary Scale (a measure of semantic impairment) for the temporal lobe variant FTLD group. Within the latter group of patients, the general neuropsychological and experimental test performance of subjects with predominantly left-sided versus predominantly right-sided temporal lobe atrophy was compared: we report differences in means with 95% Wald-type bootstrap confidence intervals (2000 replicates).

## Neuroimaging protocol and analysis

### Brain image acquisition

For 18 patients with Alzheimer's disease and 11 patients with temporal lobe variant FTLD, T<sub>1</sub>-weighted volumetric magnetic resonance images were acquired on a Siemens Trio TIM 3T scanner (Siemens Medical Systems). Scans were acquired using a 3D magnetization prepared rapid gradient echo (MP-RAGE) sequence producing 208 contiguous 1.1 mm thick sagittal slices with 28 cm field of view and a 256 × 256 acquisition matrix, giving approximately isotropic 1.1 mm cubic voxels.

### Voxel-based morphometry analyses

Magnetic resonance brain images were processed using MATLAB 7.2 (The MathWorks Inc.) and SPM8 software (Statistical Parametric Mapping, Version 8; <http://www.fil.ion.ucl.ac.uk/spm>) with default settings for all parameters; normalization was performed using the DARTEL toolbox (Ashburner, 2007). Further details of preprocessing steps are provided in the Supplementary material.

For each modality in the experimental battery (voices, faces, names), associations between regional grey matter volume and subtest performance were assessed in both disease groups using linear regression models. In separate-modality design matrices, grey matter volume was modelled as a function of the experimental subtest score-by-group interaction term with group, age and total intracranial volume included as covariates. Where the interaction was found to be significant, within-group associations were investigated further. Where no significant group interaction was identified, grey matter volume was modelled as a function of experimental subtest score in both disease groups, with covariates of group, age and total intracranial volume. The latter was measured outside statistical parametric mapping using a previously described procedure (Whitwell *et al.*, 2001). In addition to these separate-modality analyses, joint combined-modalities models were used to assess the independent partial associations of voice, face and name modalities for the familiarity subtest and the identification subtest and partial associations of voice and face modalities for the cross-modal recognition subtest. For each subtest, F tests were used to assess grey matter associations with performance for each modality (adjusting for the others) and conjointly across modalities.

An explicit analysis mask was used to exclude any voxels for which >20% of the images had an intensity value of <0.1 (Ridgway *et al.*, 2009). Grey matter associations were assessed over the whole brain and within the regions of interest specified by our prior anatomical hypotheses; small volumes covering temporal and parietal lobe regions specified in our prior anatomical hypotheses were created manually in MRICron® (<http://www.cabiatl.com/mricron/mricron/index.html>) from a study-specific template image (further details in Supplementary material). A voxel-wise statistical threshold  $P < 0.05$  family-wise-error-corrected for multiple comparisons was applied in all analyses (a global  $P < 0.05$  family-wise error-corrected threshold was applied in the combined-modalities conjunction analysis). Statistical parametric maps were displayed as overlays on the study-specific template. A voxel-wise exclusive masking procedure was applied to display grey matter areas associated with voice processing performance but not with performance in other modalities. In order to report coordinates of local maxima in the standard stereotactic Montreal Neurological Institute space, the grey matter segment of the final DARTEL template was affine registered to the *a priori* grey matter tissue probability map in statistical parametric maps, and the DARTEL coordinates were transformed using the estimated affine mapping to Montreal Neurological Institute space.

## Results

### General neuropsychological data

Results of the general neuropsychological assessment are presented in Table 2. Relative to the healthy control group both the temporal lobe variant FTL D group and the Alzheimer's disease group showed reduced verbal and performance IQ and deficits of executive function and cognitive speed, recognition memory for words and faces, naming and calculation; the Alzheimer's disease group showed additional deficits of visual object perception and auditory verbal working memory. The temporal lobe variant FTL D group had lower verbal and reading IQ than the Alzheimer's disease group and performed significantly worse than the Alzheimer's disease group on tests of naming, tests of semantic knowledge (verbal comprehension and London landmark identification tests), reading and recognition memory for faces; while the Alzheimer's disease group performed significantly worse than the temporal lobe variant FTL D group on tests of non-verbal reasoning and recognition memory for words. Considered together, these patterns of performance support the clinical and neuroanatomical classification for each disease group.

### Peripheral hearing data

On screening assessment of peripheral hearing, increasing age was associated (as anticipated) with a significant increase in mean detection threshold at the three highest frequencies tested (2, 3 and 4 kHz). Relative to the healthy control group, there was a significant difference ( $P < 0.05$ ) in mean detection thresholds at 0.5 kHz for both patient groups (adjusted differences in means from controls: Temporal lobe variant FTL D group = 7.2 dB, Alzheimer's disease group = 4.7 dB); these threshold elevations were small and unlikely to be clinically relevant. No significant differences were shown by either patient group with respect to controls at any other frequency tested, and no significant differences were

observed between temporal lobe variant FTL D and Alzheimer's disease groups at any frequency.

### Experimental behavioural data

Here we report the findings from the group analyses on tests in the experimental battery. Details of individual subject data are provided in the Supplementary material.

### Perceptual processing of voice attributes

Results for the patient and healthy control groups on early perceptual and apperceptive subtests for each modality are summarized in Table 3 and further details are provided in Supplementary Table 2.

On the vocal gender subtest, the Alzheimer's disease group performed significantly worse ( $P < 0.05$ ) than the healthy control group, however this difference was driven by a subgroup of four patients with Alzheimer's disease (the remaining patients scored at ceiling on this task). The performance of the temporal lobe variant FTL D group did not differ from healthy controls, however all subjects in the temporal lobe variant FTL D and control groups performed at ceiling on this subtest. On the vocal size subtest, there were no significant group performance differences and a large range of scores in all three groups.

On both the 'easy' and the 'difficult' speaker discrimination subtests, the Alzheimer's disease group performed significantly worse ( $P < 0.05$ ) than healthy controls. There were no significant performance differences between the temporal lobe variant FTL D group and healthy controls or between the two patient groups.

### Recognition of voices, faces and names

Results for the patient and healthy control groups on semantic subtests for each modality are summarized in Table 4 and further details are provided in Supplementary Tables 2, 3 and 4.

On all semantic subtests, both the temporal lobe variant FTL D group ( $P < 0.001$ ) and the Alzheimer's disease group ( $P < 0.05$ )

**Table 3** Behavioural data: perceptual and apperceptive processing of voices and faces

Subtest (max score)	Temporal lobe variant FTL D <i>n</i> = 14		Alzheimer's disease <i>n</i> = 22		Healthy controls <i>n</i> = 35		Temporal lobe variant FTL D–Alzheimer's disease  Difference in means (95% confidence interval)
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range	
Voice perception							
Size perception (/20)	16.7 (2.8)	11–20	17.4 (2.1)	12–20	17.1 (2.9)	9–20	–0.7 (–2.4, 1.0)
Gender perception (/24)	24.0 (0.0)	24–24	23.7 (0.6)*	22–24	24 (0.0)	24–24	0.3 (0.01, 0.5) <sup>†</sup>
Easy speaker discrimination (/28)	24.7 (1.6)	22–27	24.1 (3.2)*	15–28	25.6 (1.5)	21–28	0.6 (–1.0, 2.2)
Difficult speaker discrimination (/12)	9.2 (1.2)	7–11	8.8 (1.7)**	6–12	9.9 (1.4)	7–12	0.4 (–0.5, 1.4)
Face perception							
Benton facial recognition test (/56)	42.8 (4.0)**	37–50	42.2 (5.8)**	32–52	48.0 (3.2)	42–56	0.7 (–2.5, 3.8)

\*Significantly worse than controls ( $P < 0.05$ ); \*\*significantly worse than controls ( $P < 0.01$ ); <sup>†</sup>Alzheimer's disease group significantly worse than temporal lobe variant FTL D group ( $P < 0.05$ ).

**Table 4** Behavioural data: semantic processing of voices, faces and names

Subtest (max score)	Temporal lobe variant FTLD <i>n</i> = 14		Alzheimer's disease <i>n</i> = 22		Healthy controls <i>n</i> = 35		Temporal lobe variant FTLD–Alzheimer's disease Difference in means (95% confidence interval)
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range	
Familiarity							
Voice (/48)	27.5 (4.8)**	22–41	34.4 (5.5)**	24–45	41.5 (2.9)	35–46	–6.9 <sup>‡</sup> (–10.2, –3.7)
Face (/48)	34.6 (7.1)**	19–45	39.0 (7.1)**	26–48	46.6 (1.7)	41–48	–4.4 (–9.2, 0.3)
Name (/48)	34.6 (7.2)**	24–47	44.8 (3.2)*	33–48	46.6 (1.8)	42–48	–10.1 <sup>‡</sup> (–14.1, –6.2)
Naming							
Voice (/24)	0.6 (1.6)**	0–6	3.2 (3.4)**	0–11	17.4 (3.9)	9–24	–2.6 <sup>†</sup> (–4.3, –0.9)
Face (/24)	2.2 (3.8)**	0–14	7.0 (5.7)**	0–19	21.6 (2.6)	15–24	–4.7 <sup>†</sup> (–7.8, –1.7)
Identification							
Voice (/24)	2.6 (5.1)**	0–19	10.3 (7.0)**	0–22	19.5 (3.1)	14–24	–7.7 <sup>‡</sup> (–11.5, –3.8)
Face (/24)	7.7 (7.5)**	0–22	17.4 (6.0)**	2–24	23.6 (0.8)	21–24	–9.7 <sup>‡</sup> (–14.2, –5.1)
Name (/24)	7.2 (7.3)**	0–20	19.6 (4.1)**	10–24	23.9 (0.3)	23–24	–12.4 <sup>‡</sup> (–16.6, –8.2)
Cross-modal matching							
Voice (/24)	6.4 (7.1)** <sup>a</sup>	1–24	17.4 (6.4)**	5–24	23.6 (0.8)	21–24	–11.2 <sup>‡</sup> (–15.7, –6.7)
Face (/24)	10.1 (7.6)**	2–23	19.6 (5.0)**	6–24	24.0 (0.0)	24–24	–9.5 <sup>‡</sup> (–13.9, –5.0)

a One patient with temporal lobe variant FTLT scored 1/13 on the first 13 items on the task and declined to continue the test; his results were included in the analysis as a chance score (3/24).

\*Significantly worse than controls ( $P < 0.05$ ); \*\*significantly worse than controls ( $P < 0.001$ ); <sup>†</sup>temporal lobe variant FTLT group significantly worse than Alzheimer's disease group ( $P < 0.01$ ); <sup>‡</sup>temporal lobe variant FTLT group significantly worse than Alzheimer's disease group ( $P < 0.001$ ).

performed significantly worse than the healthy control group. For both disease groups and also for the healthy control group, mean absolute scores across semantic subtests were lower for voice recognition than for recognition in the other modalities. The temporal lobe variant FTLT group performed significantly worse than the Alzheimer's disease group ( $P < 0.01$ ) on all familiarity subtests apart from face familiarity [for which there was a trend ( $P = 0.07$ ) to worse performance], on all identification subtests in each modality, on the cross-modal subtests and on voice and face naming.

There was a significant interaction between group and modality for all subtests: familiarity ( $P < 0.001$ ), identification ( $P < 0.05$ ), cross-modal recognition ( $P < 0.01$ ) and naming ( $P < 0.01$ ). The temporal lobe variant FTLT group showed a significantly larger ( $P < 0.05$ ) decrease in score compared to the Alzheimer's disease group for identification in the name modality compared with the voice modality; the temporal lobe variant FTLT–Alzheimer's disease performance discrepancy did not differ significantly between modalities for the other subtests. In general, significant performance correlations were observed between voice recognition subtests in the Alzheimer's disease group though less consistently in the temporal lobe variant FTLT group; while voice and face identification performance was correlated in both disease groups (Supplementary Table 3). Voice identification performance was not closely correlated with apperceptive performance in either disease group (Supplementary Table 2). Voice identification and familiarity performance were not correlated with disease severity measures in either disease group (Supplementary Table 4).

## Right versus left temporal lobe damage in temporal lobe variant FTLT

Temporal lobe variant FTLT subgroups with predominantly left-sided ( $n = 9$ ) versus predominantly right-sided ( $n = 4$ ) temporal lobe atrophy did not show significant performance differences on any of the voice processing subtests. Further details are provided in the Supplementary material.

## Neuroanatomical data

### Interactions between disease group and performance

No significant grey matter associations were identified for group–performance interactions for any of the experimental subtests over the whole brain volume. Restricting analyses to the pre-specified anatomical volume of interest there was a significant interaction between group and performance on the 'easy' speaker discrimination task in the right parahippocampal gyrus (local maximum Montreal Neurological Institute coordinates: 35–51–6; cluster size 123 voxels,  $P < 0.05$  after family-wise error correction). Voice discrimination performance in the Alzheimer's disease group (but not the temporal lobe variant FTLT group) was positively associated with grey matter in right inferior parietal cortex ( $P < 0.05$  after family-wise error correction over the pre-specified small volume of interest; Table 5); additional associations of voice discrimination were present in right parahippocampal gyrus and left inferior parietal cortex at an uncorrected threshold ( $P < 0.001$  over the whole brain volume; Fig. 1).

**Table 5** Voxel-based morphometry data: neuroanatomical associations of behavioural performance

Task	Neuroanatomical associations						
	Side	Area	Cluster size (voxels)	Local maxima			
				Coordinates (mm)			Z-score
Apperceptive Voice discrimination <sup>a</sup>	R	Inferior parietal cortex	838	56	–50	42	4.13
Familiarity Face	L	Temporal pole	678	–55	5	–42	4.29
		Anterior middle temporal gyrus		–62	2	–30	4.11
	R	Anterior fusiform gyrus	250	28	4	–53	4.20
Voice, face and name <sup>b</sup>	R	Anterior fusiform gyrus	1203	35	–20	–38	4.39
Identification Voice	R	Temporal pole	3829	28	20	–42	4.55
		Hippocampus		35	–10	–21	4.26
		Entorhinal cortex		32	1	–36	4.21
		Amygdala		32	–7	–28	4.18
Face	R	Temporal pole*	558	25	18	–45	4.18
		Anterior fusiform gyrus*		30	8	–50	5.40
	L	Temporal pole	481	–47	8	–47	5.39
Name	R	Anterior fusiform gyrus*	18	32	–16	–42	4.76
		Temporal pole	2780	25	18	–46	4.39
	L	Temporal pole	942	–47	3	–45	4.15
Voice, face and name <sup>b</sup>	R	Temporal pole*	1861	25	18	–45	4.76
		Anterior fusiform gyrus		32	–17	–41	4.58
Cross-modal matching Voice	R	Temporal pole*	16	24	18	–42	4.90
		Anterior fusiform gyrus*	3098	32	–17	–41	4.49
		Entorhinal cortex		32	1	–40	4.32
Face	R	Temporal pole*	2712	25	18	–46	4.58
		Anterior fusiform gyrus		32	–15	–43	4.20
Voice and face <sup>b</sup>	R	Temporal pole	1159	24	18	–42	4.47
		Anterior fusiform gyrus		32	–17	–41	4.20

Results for voice discrimination were derived from the Alzheimer's disease group only; all other results were derived across the temporal lobe variant FTL and Alzheimer's disease groups. All clusters of size >10 voxels are presented.

a 'Easy' version of speaker discrimination task (see text).

b Results based on combined-modalities analyses; other results based on separate-modality analyses (see text).

\*Areas with local maxima exceeding voxel-wise significance threshold  $P < 0.05$  after family-wise error correction over the whole brain; other local maxima after correction over the pre-specified small volume of interest (co-ordinates in Montreal Neurological Institute stereotactic space).

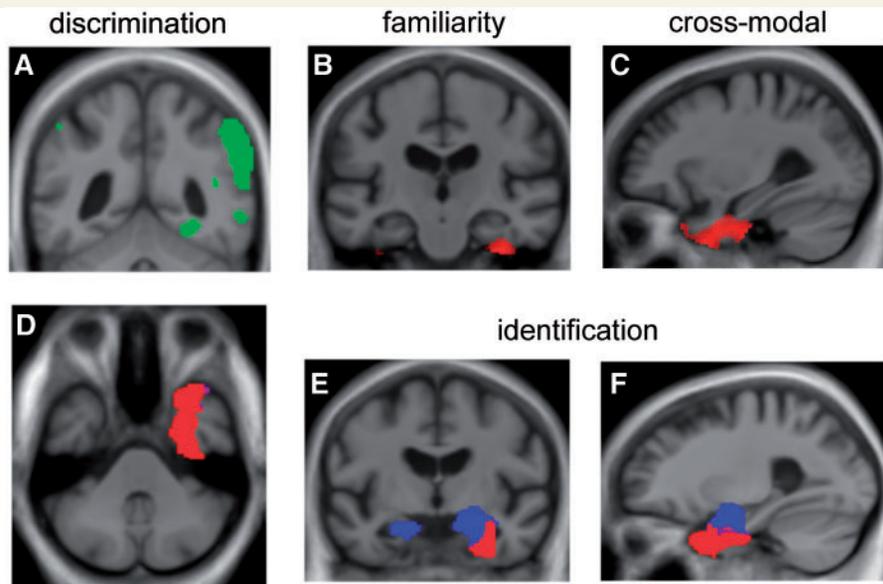
### Associations of performance across disease groups

The results of the neuroanatomical analysis across both the temporal lobe variant FTL and Alzheimer's disease groups (adjusting for group membership) are summarized in Table 5; statistical parametric maps are presented in Fig. 1.

We consider first the results of analyses for associations of experimental test performance over the whole brain volume. No significant associations of voice perceptual performance across both disease groups were identified. In the separate-modality analyses of semantic processing, cross-modal recognition of voices and identification and cross-modal recognition of faces were each positively associated with grey matter volume at the right temporal pole; in addition, cross-modal recognition of voices and identification of faces and names were each positively associated with grey matter volume in right anterior fusiform gyrus (all  $P < 0.05$  after family-wise error correction over the whole brain volume). In the combined-modalities analysis, there was a common grey matter association of voice, face and name identification at the right

temporal pole ( $P < 0.05$  after family-wise error correction over the whole brain volume), however no significant partial associations of voice, face or name identification were identified.

Restricting analyses to the pre-specified anatomical volumes of interest, a number of additional associations were identified (all  $P < 0.05$  after family-wise error correction over the relevant small volume). In the separate-modality analyses of semantic processing, across both disease groups, voice and name identification were each positively associated with grey matter at the right temporal pole; voice identification (but not face or name identification) was positively associated with grey matter in right amygdala and hippocampus, while face and name identification (but not voice identification) were each positively associated with grey matter at the left temporal pole. In the combined-modalities analysis, a common grey matter association of voice, face and name familiarity was identified in right fusiform gyrus; common grey matter associations of voice and face cross-modal recognition were identified in right temporal pole and anterior fusiform



**Figure 1** Statistical parametric maps of grey matter volume associated with voice processing performance. Statistical parametric maps show grey matter associations of experimental test performance (Table 5). (A) Speaker discrimination (Alzheimer's disease group only), (B) voice familiarity, (C) cross-modal matching of familiar voices and faces and (D–F) voice identification (all for temporal lobe variant FTLD and Alzheimer's disease groups combined). The colour code indicates areas associated with apperceptive processing of voices (green), semantic processing of voices as well as faces and names (red) and areas associated with identification of voices but not faces or names after exclusive masking (blue). Statistical parametric maps are presented on sections of the mean normalized  $T_1$ -weighted structural brain image in DARTTEL space. Coronal (A, B, E), axial (D) and sagittal (C, F) sections are shown, targeting inferior parietal lobes (A), anterior and inferior temporal lobes (B–F). The sagittal sections are derived from the right hemisphere and the right hemisphere is shown on the right in all other sections. All statistical parametric maps are based on regions for which grey matter associations were significant ( $P < 0.05$ ) after correction for multiple comparisons over the pre-specified anatomical small volume (Table 5); statistical parametric maps are thresholded at  $P < 0.001$  uncorrected for display purposes.

gyrus. No significant partial associations were identified for voice, face or name familiarity or for cross-modal recognition of voices or faces. No significant grey matter associations of voice or face naming performance were identified.

## Discussion

Here we have described behavioural and neuroanatomical signatures of voice processing deficits in FTLD and Alzheimer's disease. The key findings from the study are two-fold. In behavioural terms, impairments of voice recognition are significant in these canonical dementia syndromes but particularly severe in temporal lobe variant FTLD. In neuroanatomical terms, impaired voice recognition is associated with grey matter loss in the right anterior temporal lobe, suggesting that this region is critical for voice recognition as well as other modalities of person knowledge.

On all behavioural measures of voice recognition (familiarity, identification, naming and cross-modal identity matching), both disease groups showed deficits relative to healthy controls and the temporal lobe variant FTLD group performed significantly worse than the Alzheimer's disease group. Despite substantial individual variation in performance, both disease groups showed a clear overall trend to conjoined deficits of voice recognition with deficits of other modalities of person knowledge (in particular,

recognition of faces). The largest performance discrepancy between the disease groups occurred for the processing of personal names. On measures of perceptual processing of voices, the Alzheimer's disease group (but not the temporal lobe variant FTLD group) showed deficits relative to healthy controls, suggesting that impairments of voice perception may show relative specificity for Alzheimer's disease.

The neuroanatomical analysis showed that recognition measures for voice, face and name processing were associated across both disease groups with grey matter volume in right temporal pole and anterior fusiform gyrus. Similar regions have been implicated in the processing of familiar voices by the healthy brain (Belin *et al.*, 2000, 2002; Nakamura *et al.*, 2001; Shah *et al.*, 2001; Belin and Zatorre, 2003; von Kriegstein *et al.*, 2003, 2005; von Kriegstein and Giraud, 2004; Andics *et al.*, 2010). These same areas are sites of heavy disease involvement in temporal lobe variant FTLD, consistent with the more severe person recognition deficits in this group compared with the Alzheimer's disease group. However, it is unlikely the profile of anatomical associations observed was driven entirely by the temporal lobe variant FTLD group, since there was no evidence of significantly different grey matter associations of semantic test performance between the disease groups. We do not wish to over-emphasize this apparent anatomical convergence. Whereas Alzheimer's disease is pathologically homogeneous, temporal lobe variant FTLD is likely to

be pathologically heterogeneous (Josephs *et al.*, 2011): a shared macro-anatomical substrate may be underpinned by distinct patterns of cellular involvement and correspondingly distinct pathophysiological mechanisms. Moreover it remains possible that there is a more fine-grained segregation of processing for different modalities within the relatively large cortical areas identified here using voxel-based morphometry (Olson *et al.*, 2007).

Certain mesial temporal lobe structures (amygdala, hippocampus and entorhinal cortex) showed an association with voice identification but not other modalities of person knowledge at the prescribed threshold (Fig. 1); the mesial temporal lobe has been previously implicated in familiar voice processing by healthy subjects (Nakamura *et al.*, 2001; von Kriegstein and Giraud, 2004; von Kriegstein *et al.*, 2005; Andics *et al.*, 2010) and this region may be involved in processing sensory object information (Lee *et al.*, 2005) and particularly in tracking information in sound (for example, familiar musical melodies): (Samson and Zatorre, 1992; Watanabe *et al.*, 2008). However, any apparent modality-specificity here should be interpreted with caution; no independent associations of recognition in a particular modality emerged when modalities were assessed together.

Both the behavioural and neuroanatomical findings here are consistent with a growing body of evidence implicating the anterior temporal lobe in multi-modal processing of semantic knowledge and more particularly, person knowledge (Bozeat *et al.*, 2000; Coccia *et al.*, 2004; Luzzi *et al.*, 2007; Rami *et al.*, 2007; Lambon Ralph and Patterson, 2008). Previous functional imaging and lesion evidence suggests that the temporal pole and anterior fusiform participate in a cooperative network mediating different aspects of semantic processing (Ellis *et al.*, 1989; Tranel *et al.*, 1997; Papagno and Capitani, 1998; Grabowski *et al.*, 2001; Thompson *et al.*, 2004; von Kriegstein *et al.*, 2005; Tranel, 2006; Mion *et al.*, 2010). While we cannot specify the precise role of anterior fusiform and temporal pole in the present study, grey matter volume in anterior fusiform was associated with performance on familiarity and cross-modal judgements across modalities, consistent with the operation of person identity nodes in this area (Belin *et al.*, 2004). We do not argue that processing voice familiarity is a purely semantic capacity: analogously with familiarity for other kinds of stimuli, voice familiarity may also have perceptual, affective and executive dimensions (Gainotti, 2007; Lucchelli and Spinnler, 2008). Taken together, our findings support the existence of multi-modal deficits of person knowledge in FTLD and Alzheimer's disease. Rather than a purely amodal or fully multi-modal organization, the data suggest that verbal (name) and non-verbal (voice, face) modalities of person knowledge may be partially differentiated, whereas modalities of non-verbal person knowledge are more closely aligned (Warrington, 1979; Snowden *et al.*, 2004). The present data (based on small case numbers) do not resolve the important issue of potentially asymmetric temporal lobe contributions to different components of person knowledge (Snowden *et al.*, 2004).

In this study, deficits of voice perception were restricted to the Alzheimer's disease group, and involved voice apperception (speaker discrimination) and encoding of one perceptual attribute (vocal gender) while sparing encoding of another attribute (vocal

size). A neuroanatomical association of apperceptive performance in the Alzheimer's disease group was identified in the right inferior parietal lobe. The present findings underline the potential for development of semantic deficits of voice recognition (and other aspects of person recognition) despite intact pre-semantic perceptual mechanisms; however, deficits of voice perception may have contributed to the development of voice recognition impairment in the Alzheimer's disease group. Functional imaging work in healthy subjects has delineated a network of cortical areas including the parietal lobe for processing voice information under non-canonical listening conditions (Bishop and Miller, 2009). We propose that inferior parietal cortex is involved in the structural representation of voices (Bruce and Young, 1986; Burton *et al.*, 1990; Belin *et al.*, 2004), perhaps by holding voice information online in working memory for comparison with incoming alternative auditory 'views' of the speaker (e.g. the same voice speaking different phonemes).

This study suggests clear directions for future work. It has been proposed that modality-specific deficits of person knowledge become generalized with the evolution of neurodegenerative disease (Evans *et al.*, 1995; Gentileschi *et al.*, 1999; Gainotti *et al.*, 2003, 2008): the present study suggests that the profile of development of deficits may hold information about the organization of processing within and between modalities. This issue will only be addressed by longitudinal studies based on a systematic analysis of different levels of processing and comparing modalities and disease groups. Whereas semantic processing of voices is relatively easily investigated by adapting standard neuropsychological techniques, a detailed understanding of voice perception and its disorders will require the design of customized stimuli that allow particular vocal attributes to be isolated and manipulated. Finally, there is a need for correlation of voice processing measures with structural and functional anatomical data and with tissue histopathology in a range of neurodegenerative diseases.

## Acknowledgements

We are grateful to all the subjects for their participation. We thank Dr Doris-Eva Bamiou for assistance with audiometric assessments, Professor Roy Patterson and Dr Martin Vestergaard for assistance in preparing vocal size stimuli; Dr John Stevens for assistance interpreting brain MRIs; and Professor EK Warrington for helpful discussion.

## Funding

This work was undertaken at University College London Hospitals/ University College London who received a proportion of funding from the Department of Health's National Institute for Health Research Biomedical Research Centres funding scheme. The Dementia Research Centre is an Alzheimer's Research UK Co-ordinating Centre. Wellcome Trust; UK Medical Research Council; Alzheimer's Research UK Senior Research Fellowship (to S.J.C.); Wellcome Trust Senior Clinical Fellowship (to J.D.W.).

## Supplementary material

Supplementary material is available at *Brain* online.

## References

- Andics A, McQueen JM, Petersson KM, Gal V, Rudas G, Vidnyanszky Z. Neural mechanisms for voice recognition. *Neuroimage* 2010; 52: 1528–40.
- Ashburner J. A fast diffeomorphic image registration algorithm. *Neuroimage* 2007; 38: 95–113.
- Ashburner J, Friston KJ. Voxel-based morphometry—the methods. *Neuroimage* 2000; 11: 805–21.
- Belin P, Zatorre RJ. Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport* 2003; 14: 2105–9.
- Belin P, Zatorre RJ, Ahad P. Human temporal-lobe response to vocal sounds. *Brain Res Cogn Brain Res* 2002; 13: 17–26.
- Belin P, Fecteau S, Bedard C. Thinking the voice: neural correlates of voice perception. *Trends Cogn Sci* 2004; 8: 129–35.
- Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. Voice-selective areas in human auditory cortex. *Nature* 2000; 403: 309–12.
- Benton AL, Hamsher KS, Varney N, Spreen O. Contributions to neuropsychological assessment: a clinical manual. Oxford: Oxford University Press; 1989.
- Bishop CW, Miller LM. A multisensory cortical network for understanding speech in noise. *J Cogn Neurosci* 2009; 21: 1790–805.
- Bozeat S, Lambon Ralph MA, Patterson K, Garrard P, Hodges JR. Non-verbal semantic impairment in semantic dementia. *Neuropsychologia* 2000; 38: 1207–15.
- Brambati SM, Rankin KP, Narvid J, Seeley WW, Dean D, Rosen HJ, et al. Atrophy progression in semantic dementia with asymmetric temporal involvement: a tensor-based morphometry study. *Neurobiol Aging* 2009; 30: 103–11.
- Bruce V, Young A. Understanding face recognition. *Br J Psychol* 1986; 77: 305–27.
- Burton AM, Bruce V, Johnston RA. Understanding face recognition with an interactive activation model. *Br J Psychol* 1990; 81: 361–80.
- Campanella S, Belin P. Integrating face and voice in person perception. *Trends Cogn Sci* 2007; 11: 535–43.
- Chan D, Anderson V, Pijnenburg Y, Whitwell J, Barnes J, Schill R, et al. The clinical profile of right temporal lobe atrophy. *Brain* 2009; 132: 1287–98.
- Coccia M, Bartolini M, Luzzi S, Provinciali L, Ralph MA. Semantic memory is an amodal, dynamic system: evidence from the interaction of naming and object use in semantic dementia. *Cogn Neuropsychol* 2004; 21: 513–27.
- Crutch SJ, Warrington EK. The semantic organisation of proper nouns: the case of people and brand names. *Neuropsychologia* 2004; 42: 584–96.
- Delis DC, Kaplan E, Kramer JH. Delis-Kaplan executive function system. San Antonio: The Psychological Corporation; 2001.
- Dubois B, Feldman HH, Jacova C, Dekosky ST, Barberger-Gateau P, Cummings J, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurol* 2007; 6: 734–46.
- Ellis AW, Young AW, Critchley EM. Loss of memory for people following temporal lobe damage. *Brain* 1989; 112: 1469–83.
- Ellis HD, Jones DM, Mosdell N. Intra- and inter-modal repetition priming of familiar faces and voices. *Br J Psychol* 1997; 88: 143–56.
- Evans JJ, Higgs AJ, Antoun N, Hodges JR. Progressive prosopagnosia associated with selective right temporal lobe atrophy. A new syndrome? *Brain* 1995; 118: 1–13.
- Gainotti G. Face familiarity feelings, the right temporal lobe and the possible underlying neural mechanisms. *Brain Res Rev* 2007; 56: 214–35.
- Gainotti G, Barbier A, Marra C. Slowly progressive defect in recognition of familiar people in a patient with right anterior temporal atrophy. *Brain* 2003; 126: 792–803.
- Gainotti G, Ferraccioli M, Quaranta D, Marra C. Cross-modal recognition disorders for persons and other unique entities in a patient with right fronto-temporal degeneration. *Cortex* 2008; 44: 238–48.
- Garrido L, Eisner F, McGettigan C, Stewart L, Sauter D, Hanley JR, et al. Developmental phonagnosia: a selective deficit of vocal identity recognition. *Neuropsychologia* 2009; 47: 123–31.
- Gentileschi V, Sperber S, Spinnler H. Progressive defective recognition of familiar people. *Neurocase* 1999; 5: 407–24.
- Gentileschi V, Sperber S, Spinnler H. Crossmodal agnosia for familiar people as a consequence of right infero-polar temporal atrophy. *Cogn Neuropsychol* 2001; 18: 439–63.
- Goll JC, Crutch SJ, Warren JD. Central auditory disorders: toward a neuropsychology of auditory objects. *Curr Opin Neurol* 2010; 23: 617–27.
- Grabowski TJ, Damasio H, Tranel D, Ponto LL, Hichwa RD, Damasio AR. A role for left temporal pole in the retrieval of words for unique entities. *Hum Brain Mapp* 2001; 13: 199–212.
- Griffiths TD, Warren JD. What is an auditory object? *Nat Rev Neurosci* 2004; 5: 887–92.
- Hailstone JC, Crutch SJ, Vestergaard MD, Patterson RD, Warren JD. Progressive associative phonagnosia: a neuropsychological analysis. *Neuropsychologia* 2010; 48: 1104–14.
- Hanley JR, Damjanovic L. It is more difficult to retrieve a familiar person's name and occupation from their voice than from their blurred face. *Memory* 2009; 17: 830–39.
- Hanley JR, Young AW, Pearson NA. Defective recognition of familiar people. *Cogn Neuropsychol* 1989; 6: 179–210.
- Imazumi S, Mori K, Kiritani S, Kawashima R, Sugiura M, Fukuda H, et al. Vocal identification of speaker and emotion activates different brain regions. *Neuroreport* 1997; 8: 2809–12.
- Ives DT, Smith DRR, Patterson RD. Discrimination of speaker size from syllable phrases. *J Acoust Soc Am* 2005; 118: 3816–22.
- Josephs KA, Whitwell JL, Vemuri P, Senjem ML, Boeve BF, Knopman DS, et al. The anatomic correlate of prosopagnosia in semantic dementia. *Neurology* 2008; 71: 1628–33.
- Joubert S, Felician O, Barbeau E, Sontheimer A, Barton JJ, Ceccaldi M, et al. Impaired configurational processing in a case of progressive prosopagnosia associated with predominant right temporal lobe atrophy. *Brain* 2003; 126: 2537–50.
- Joubert S, Felician O, Barbeau E, Sontheimer A, Guedj E, Ceccaldi M, et al. Progressive prosopagnosia: clinical and neuroimaging results. *Neurology* 2004; 63: 1962–5.
- Kawahara H, Irino T. Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation. In: Divenyi PL, editor. *Speech separation by humans and machines*. Massachusetts: Kluwer Academic; 2004. p. 167–80.
- Lambon Ralph MA, Patterson K. Generalization and differentiation in semantic memory: insights from semantic dementia. *Ann N Y Acad Sci* 2008; 1124: 61–76.
- Lang CJ, Kneidl O, Hielscher-Fastabend M, Heckmann JG. Voice recognition in aphasic and non-aphasic stroke patients. *J Neurol* 2009; 256: 1303–6.
- Lee AC, Buckley MJ, Pegman SJ, Spiers H, Schill VL, Gaffan D, et al. Specialization in the medial temporal lobe for processing of objects and scenes. *Hippocampus* 2005; 15: 782–97.
- Lucchelli F, Spinnler H. A reappraisal of person recognition and identification. *Cortex* 2008; 44: 230–7.
- Luzzi S, Snowden JS, Neary D, Coccia M, Provinciali L, Lambon Ralph MA. Distinct patterns of olfactory impairment in Alzheimer's disease, semantic dementia, frontotemporal dementia, and corticobasal degeneration. *Neuropsychologia* 2007; 45: 1823–31.
- McCarthy RA, Warrington EK. Actors but not scripts: the dissociation of people and events in retrograde amnesia. *Neuropsychologia* 1992; 30: 633–44.

- Mion M, Patterson K, Acosta-Cabrero J, Pengas G, Izquierdo-Garcia D, Hong YT, et al. What the left and right anterior fusiform gyri tell us about semantic memory. *Brain* 2010; 133: 3256–68.
- Nakamura K, Kawashima R, Sugiura M, Kato T, Nakamura A, Hatano K, et al. Neural substrates for recognition of familiar voices: a PET study. *Neuropsychologia* 2001; 39: 1047–54.
- Neary D, Snowden JS, Gustafson L, Passant U, Stuss D, Black S, et al. Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. *Neurology* 1998; 51: 1546–54.
- Nelson HE. National Adult Reading Test (NART): test manual. Windsor: NFER Nelson; 1982.
- Neuner F, Schweinberger SR. Neuropsychological impairments in the recognition of faces, voices, and personal names. *Brain Cogn* 2000; 44: 342–66.
- Olson IR, Plotzker A, Ezzyat Y. The enigmatic temporal pole: a review of findings on social and emotional processing. *Brain* 2007; 130: 1718–31.
- Papagno C, Capitani E. Proper name anomia: a case with sparing of the first-letter knowledge. *Neuropsychologia* 1998; 36: 669–79.
- Rami L, Loy CT, Hailstone J, Warren JD. Odour identification in fronto-temporal lobar degeneration. *J Neurol* 2007; 254: 431–5.
- Ridgway GR, Omar R, Ourselin S, Hill DL, Warren JD, Fox NC. Issues with threshold masking in voxel-based morphometry of atrophied brains. *Neuroimage* 2009; 44: 99–111.
- Samson S, Zatorre RJ. Learning and retention of melodic and verbal information after unilateral temporal lobectomy. *Neuropsychologia* 1992; 30: 815–26.
- Schonell F, Goodacre E. The psychology and teaching of reading. London: Oliver Boyd; 1971.
- Schweinberger SR, Herholz A, Stief V. Auditory long-term memory: repetition priming of voice recognition. *Q J Exp Psychol* 1997a; 50: 498–517.
- Schweinberger SR, Herholz A, Sommer W. Recognizing famous voices: influence of stimulus duration and different types of retrieval cues. *J Speech Hear Res* 1997b; 40: 453–63.
- Seeley WW, Crawford RK, Zhou J, Miller BL, Greicius MD. Neurodegenerative diseases target large-scale human brain networks. *Neuron* 2009; 62: 42–52.
- Shah NJ, Marshall JC, Zafiris O, Schwab A, Zilles K, Markowitsch HJ, et al. The neural correlates of person familiarity. A functional magnetic resonance imaging study with clinical implications. *Brain* 2001; 124: 804–15.
- Snowden JS, Thompson JC, Neary D. Knowledge of famous faces and names in semantic dementia. *Brain* 2004; 127: 860–72.
- Thompson SA, Graham KS, Williams G, Patterson K, Kapur N, Hodges JR. Dissociating person-specific from general semantic knowledge: roles of the left and right temporal lobes. *Neuropsychologia* 2004; 42: 359–70.
- Tranel D. Impaired naming of unique landmarks is associated with left temporal polar damage. *Neuropsychology* 2006; 20: 1–10.
- Tranel D, Damasio H, Damasio AR. A neural basis for the retrieval of conceptual knowledge. *Neuropsychologia* 1997; 35: 1319–27.
- Van Lancker D, Kreiman J. Voice discrimination and recognition are separate abilities. *Neuropsychologia* 1987; 25: 829–34.
- Van Lancker DR, Canter GJ. Impairment of voice and face recognition in patients with hemispheric damage. *Brain Cogn* 1982; 1: 185–95.
- Van Lancker DR, Kreiman J, Cummings J. Voice perception deficits: neuroanatomical correlates of phonagnosia. *J Clin Exp Neuropsychol* 1989; 11: 665–74.
- Van Lancker DR, Cummings JL, Kreiman J, Dobkin BH. Phonagnosia: a dissociation between familiar and unfamiliar voices. *Cortex* 1988; 24: 195–209.
- von Kriegstein K, Kleinschmidt A, Giraud AL. Voice recognition and cross-modal responses to familiar speakers' voices in prosopagnosia. *Cereb Cortex* 2006; 16: 1314–22.
- von Kriegstein K, Eger E, Kleinschmidt A, Giraud AL. Modulation of neural responses to speech by directing attention to voices or verbal content. *Brain Res Cogn Brain Res* 2003; 17: 48–55.
- von Kriegstein K, Kleinschmidt A, Sterzer P, Giraud AL. Interaction of face and voice areas during speaker recognition. *J Cogn Neurosci* 2005; 17: 367–76.
- von Kriegstein KV, Giraud AL. Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *Neuroimage* 2004; 22: 948–55.
- Warren JD, Scott SK, Price CJ, Griffiths TD. Human brain mechanisms for the early analysis of voices. *Neuroimage* 2006; 31: 1389–97.
- Warrington EK. Neuropsychological evidence for multiple memory systems. *Ciba Found Symp* 1979; 69: 153–66.
- Warrington EK. Recognition memory test. Windsor: NFER-Nelson; 1984.
- Warrington EK. The graded naming test: a restandardisation. *Neuropsychol Rehabil* 1997; 7: 143–46.
- Warrington EK, James M. The visual object and space perception battery. Bury St Edmunds: Thames Valley Test Company; 1991.
- Warrington EK, McKenna P, Orpwood L. Single word comprehension: a concrete and abstract word synonym test. *Neuropsychol Rehabil* 1998; 8: 143–54.
- Watanabe T, Yagishita S, Kikyo H. Memory of music: roles of right hippocampus and left inferior frontal gyrus. *Neuroimage* 2008; 39: 483–91.
- Wechsler D. Wechsler memory scale-revised. San Antonio: The Psychological Corporation; 1987.
- Wechsler D. Wechsler abbreviated scale of intelligence. San Antonio: The Psychological Corporation; 1999.
- Whiteley AM, Warrington EK. Selective impairment of topographical memory: a single case study. *J Neurol Neurosurg Psychiatry* 1978; 41: 575–78.
- Whitwell JL, Crum WR, Watt HC, Fox NC. Normalization of cerebral volumes by use of intracranial volume: implications for longitudinal quantitative MR imaging. *AJNR Am J Neuroradiol* 2001; 22: 1483–89.