MECHANISMS OF VOICE PROCESSING IN DEMENTIA

A Thesis
presented for the degree of
Doctor of Philosophy
in Neuropsychology

by

Julia Catherine Hailstone

The Dementia Research Centre
Institute of Neurology
UCL

2012

Declaration

I, Julia Hailstone confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis. All participants gave informed consent and all experimental work was carried out with the approval of the University College London Hospitals' Research Ethics Committee, in accordance with the guidelines established by the Declaration of Helsinki.

_____

Julia C Hailstone Date

Abstract

Perception of nonverbal vocal information is essential in our daily lives. Patients with degenerative dementias commonly have difficulty with such aspects of vocal communication; however voice processing has seldom been studied in these diseases. This thesis comprises a series of linked studies of voice processing in canonical dementias: Alzheimer's disease, behavioural variant frontotemporal dementia, semantic dementia and progressive nonfluent aphasia. A series of neuropsychological tests were developed to examine perceptual and semantic stages of voice processing and to assess two aspects of accent processing: comprehension of foreign accented speech and recognition of regional and foreign accents; patient performance was referenced to healthy control subjects. Neuroanatomical associations of voice processing performance were assessed using voxel based morphometry. Following a symptom-led approach, a syndrome of progressive associative phonagnosia was characterised in two detailed case studies. Following a disease-led approach, this work was extended systematically to cohorts of patients representing the target diseases and assessing voice processing in relation to other aspects of person recognition (faces and names). This work provided evidence for separable profiles of voice processing impairment in different diseases: associative deficits were particularly severe in semantic dementia, whilst perceptual deficits showed relative specificity for Alzheimer's disease. Neuroanatomical associations were identified for voice recognition in the right temporal pole and anterior fusiform gyrus, and for voice discrimination in the right inferior parietal lobe. The final phase of this work addressed the neuropsychological and neuroanatomical basis of accent processing, as an important dimension of nonverbal vocal analysis that is not dependent on voice identity. This work provides evidence for impaired processing of accents in progressive nonfluent aphasia and Alzheimer's with neuroanatomical associations in the anterior and superior temporal lobe. The thesis contributes new information about voice processing in the degenerative dementias and furthers our understanding of the mechanisms of human voice analysis.

Division of labour for experimental work

The experimental work described in this thesis was conducted by the author in collaboration with researchers at the Dementia Research Centre and affiliated institutions. Substantial contributions made by others are detailed below, and other contributions are credited in the acknowledgements.

Chapter 3 (Study 1)

Experimental design of voice tests: author, Jason Warren, Sebastian Crutch

Experimental design of size test: author, Martin Vestergaard and Roy Patterson

Construction and piloting of tests: author

Data collection: author

Data analysis: author in consultation with Jonathan Bartlett

Writing the published paper: author and Jason Warren

Chapters 4 & 5 (Studies 2 and 3)

Experimental design: author, Jason Warren and Sebastian Crutch

Construction and piloting of tests: author

Experimental data collection: author

Background neuropsychological data collection: author, Jo Goll and Aisling Buckley

Data analysis: author in consultation with Jonathan Bartlett

VBM image processing: author in consultation with Jo Goll

VBM analyses: author in consultation with Gerard Ridgway

Writing the published papers: author and Jason Warren

Acknowledgements

Thanks must first be expressed to my supervisors: Jason who has provided an encyclopaedic knowledge of the field and has been a major source of inspiration for the work, and to Sebastian who provided neuropsychological expertise and moral support which have been invaluable to the work and to my professional development. Together they have been a tireless source of help and support and have steered my PhD to completion. I am also indebted to two important contributors to this work: Jonathan Bartlett for continual statistical advice and Ged Ridgway for masterful consultation on VBM analyses. My deepest thanks go to the patients and control participants without whom this work would not have been possible, and Professors Martin Rossor and Nick Fox for overseeing the clinical care of the patients involved.

I am grateful to Dr Doris-Eva Bamiou for assistance with audiometric assessments, Dr John Stevens for assistance in interpretation of brain images and Professor Warrington for helpful discussion. I thank members of the admin team at the Dementia Research Centre: Suzie Barker, Anne Parnell, Ayesha Khatun and Carolyn Anderson who assisted in the logistics of patient recruitment and research visits. I am grateful to Johanna Goll, Manja Lehmann and Susie Henley for their support with VBM and general comradery. I am indebted to all the DRC members and friends that been guinea pigs for pilot neuropsychological tests and/or have allowed me to record their voice for tests in this Thesis: Adam Reid, Rick Merrick, Francesca Silman, Josephine Davies, Harry Lee and Sophie Coulombeau, Laila Ahsan, Shona Clegg, Anne Parnell, Jane Douglas, Claire Bloomfield and Mary Keilty. I am also grateful to Marina Tyndall, Tessa Mellow, Katie Piwnica-Worms, Keri Sills, Kate Butler and Kerry Fernhead for recording non-British accent samples and to staff at the British Library for their help sourcing voice recordings. Lastly I would like to thank Pete, Laila and my family for their kindness and support throughout this demanding process.

This work was undertaken at UCLH/UCL who received a proportion of funding from the Department of Health's NIHR Biomedical Research Centres funding scheme. The Dementia Research Centre is an Alzheimer's Research UK Co-ordinating Centre. This work was funded by the Wellcome Trust and by the UK Medical Research Council.

Contents

Figures presented in this Thesis

Tables presented in this Thesis

Abbreviations presented in this Thesis

AD	Alzheimer's disease
ATL	anterior temporal lobe
BPVS	British picture vocabulary scale
BvFTD	behavioral variant frontotemporal dementia
CI	confidence interval
CSF	cerebrospinal fluid
DARTEL	differomorphic anatomical registration through exponentiated lie algebra (toolbox in SPM8)
dB	decibels
F_0	fundamental frequency
FAS	foreign accent syndrome
fMRI	functional magnetic resonance imaging
FRU	face recognition units
FTLD	frontotemporal lobar degeneration
FWE	family-wise-error
GMV	grey matter volume
GNT	graded naming test
Hz	hertz
IAC	interactive activation and competition network model
KHz	kilohertz
MMSE	mini mental state examination score
MNI	Montreal Neurological Institute (standard stereotactic space)
MRI	magnetic resonance imaging
NART	national adult reading test
PIN	person identity node
PNFA	progressive non-fluent aphasia
PPA	primary progressive aphasia
ROI	region of interest
SD	semantic dementia
sec	seconds
SPM	statistical parametric map
SPM8	statistical parametric mapping software version 8
STG	superior temporal gyrus
STS	superior temporal sulcus
TIV	total intracranial volume
tvFTLD	temporal variant frontotemporal lobar degeneration
VBM	voxel-based morphometry
VRU	voice recognition units
VTL	vocal tract length

1. General Introduction

1.1. Overview

The experiments designed in this thesis address the cognitive and neural mechanisms of voice processing in neurodegenerative disease. The neuropsychology of voice processing has been relatively little studied in contrast to the vast literature on face processing in the visual domain. This thesis focuses on two aspects of voice processing: recognition of familiar voices, and processing of foreign and regional accents. These are investigated neuropsychologically and using neuroimaging methodology: voxel-based morphometry (VBM) in degenerative patients with frontotemporal lobar degeneration (FTLD) and Alzheimer's disease (AD).

The ability to recognise a voice without visual cues is a common experience: we are able to identify the voices of family members over the telephone, to find voices of media personalities on the radio familiar, and to recognise a colleague's voice from behind a closed door. Impairments of voice recognition (or phonagnosia) have been studied in cases of focal lesions involving the temporal and parietal cortices (Van Lancker, Cummings, Kreiman *et al.* 1988; Van Lancker, Kreiman, and Cummings 1989) and in degenerative cases with atrophy affecting the right anterior temporal lobe (ATL) (Evans, Heggs, Antoun *et al.* 1995; Gentileschi, Sperber, and Spinnler 1999; Gentileschi, Sperber, and Spinnler 2001). Analogous to specialized visual mechanisms proposed for face processing, functional imaging studies in healthy controls have provided evidence for specialised auditory neuroanatomical substrates for voice processing in the superior temporal lobe (Warren, Scott, Price *et al.* 2006; von Kriegstein and Giraud 2004; Belin, Zatorre, Lafaille *et al.* 2000; Belin, Zatorre, and Ahad 2002) and models of voice processing hypothesize serial processing pathway from posterior to anterior superior temporal sulcus (STS) and superior temporal gyrus (STG) (Warren, Scott, Price *et al.* 2006; von Kriegstein and Giraud 2004; Belin, Fecteau, and Bedard 2004). Other studies of familiar voice recognition have found activations in extra-temporal multimodal processing regions (Imaizumi, Mori, Kiritani *et al.* 1997; Nakamura, Kawashima, Sugiura *et al.* 2001; Shah, Marshall, Zafiris *et al.* 2001; von Kriegstein, Eger, Kleinschmidt *et al.* 2003); the stages and functions of auditory and multimodal cortical regions recruited in voice processing tasks however are underspecified.

Recognition of a person's regional or foreign accent is also an ecologically valid task, speakers with different accents regularly interact with each other; within the United Kingdom accents show

great variation even within the same county (www.bl.uk/learning/langlit/sounds), and in multicultural environments such as London people are exposed to a wide variety of foreign accents. Knowledge about variation in non-native accents is likely to facilitate an understanding of the cultural and social background of a speaker, and to enhance speech comprehension. Two types of task have been used in studies of healthy volunteers: recognition of regional or foreign accents and comprehension of speech spoken in non-native accents (Adank, Evans, Stuart-Smith *et al.* 2009; Clarke and Garrett 2004; Clopper and Pisoni 2004b; Clopper and Pisoni 2007; Floccia, Butler, Goslin *et al.* 2009; Goldstone 1994; Howell, Barry, and Vinson 2006), however few neuropsychological studies of accent processing have been carried out (Dunton, Bruce, & Newton 2011). Activations during accent recognition tasks indicate that a network involving frontal, parietal and temporal cortices may be recruited (Adank, Evans, Stuart-Smith *et al.* 2009; Berman, Mandelkern, Phan *et al.* 2003) overlapping with the network of regions involved in processing of emotion and prosody. The study of voice processing in patients with relevant pathology is required to determine which regions within the network are critical to accent or voice processing more broadly.

There are strong clinical and neurobiological grounds for a systematic analysis of voice processing in cohorts of patients with degenerative dementias. Clinically, voice processing impairments are likely to be under-recognised in these populations due to the general availability of compensatory cues from face and other contextual information, yet they may constitute a significant and disabling symptom especially in situations where such additional cues are reduced or unavailable. Disease-specific deficits (or relative preservation) of voice processing could potentially assist diagnosis of particular dementias. Neurobiologically, these diseases offer a perspective on voice processing that is complementary both to studies in normal subjects and in patients with focal brain lesions: in contrast to focal lesion studies where damage occurs as a result of random stochastic or physical factors (such as blood supply), neurodegenerative diseases disrupt functional cortical networks (Seeley, Crawford, Zhou *et al.* 2009). Anatomical degeneration of functional neural networks in dementia (Seeley *et al.*2009; Sonty, Mesulam, Weintraub *et al.* 2007; Young, Newcombe, de Haan *et al.* 1993) and systematic neuropsychological assessment of the breakdown of voice processing would potentially enable identification of critical nodes in functional and anatomical cerebral networks, and inform neurocognitive models of voice processing.

Voice processing in this thesis will be investigated in four syndromes with differing patterns of cortical atrophy: semantic dementia (SD), behavioural variant FTLD (bvFTD), Progressive Non-

fluent Aphasia (PNFA) and typical Alzheimer's disease (AD). These diseases affect areas in the temporal, parietal and frontal lobes that have been implicated in neuropsychological and neuroimaging studies of voice processing (Belin, Zatorre, & Ahad 2002; Imaizumi, Mori, Kiritani *et al.* 1997; Nakamura, Kawashima, Sugiura *et al.* 2001; von Kriegstein, Kleinschmidt, & Giraud 2006; Warren, Scott, Price *et al.* 2006) but with differing distributed patterns atrophy. In particular, the brunt of tissue damage in AD and several diseases in the FTLD spectrum initially falls on the temporal lobes, which are likely to contain mechanisms integral for voice analysis and recognition (Belin, Fecteau, & Bedard 2004). In PNFA and also in AD, atrophy impinges on cortical regions that are implicated more generally in complex sound processing such as posterior temporal auditory association cortex, whereas in SD multimodal regions are implicated. Voice processing impairments have rarely been systematically tested in these syndromes although emerging evidence suggests that all four groups may have vocal or more general auditory processing disorders. Patterns of voice processing performance and the neuroanatomical correlates of any impairment will be compared between disease groups and related to the neuropsychological and clinical profiles of each syndrome in order to improve understanding of the nonverbal symptoms these patients experience.

In the following chapter the neurology of voice and accent processing in healthy individuals and in patients with focal brain lesions are reviewed. A model of vocal processing is then outlined and the rationale for the study of voice processing in neurodegenerative disease considered.

1.2. Background: perception of paralinguistic information in voices

1.2.1. Segregation of linguistic and paralinguistic processing streams

The human voice is the acoustical signature of our species. Recognition and interpretation of the vocal signals across different contexts and listening conditions play a central role in nonverbal social communication, which has a long evolutionary history preceding language by millions of years (Fitch 2000). Not only is the voice the carrier of speech but it is an "auditory face" rich in acoustical features which communicate a wealth of information about the speaker: their identity, sex, age, social background and geographical origins (communicated via their regional accent). Paralinguistic information in the voice conveys stable characteristics of the speaker, but also serves as an important mode of communication about an individual's emotional state, mood or attitudes. Vocal information may combine with visual cues from faces in facilitating non-verbal

communication, but may also be particularly useful when the face is not visible, such as when communicating in the dark, over longer distances, or over the telephone.

Speaker characteristics and affective cues are communicated via combinations of qualities of the sounds we produce: characteristics which include the pitch, timbre (or vocal quality), regional accent, intonation (variation in pitch which is not used to distinguish words) and articulation. Characteristic patterns in these features contribute to a consistent and stable representation of a person (Belin *et al.*2004; Karpf 2006) and like face recognition allow us to recognize a person's identity from their voice. Nonverbal vocal communication is not specific to humans: non-human primates also orient to conspecific vocalizations and recognize other individuals from their voice (Ghazanfar, Turesson, Maier *et al.* 2007; Masataka 1985; Rendall 2003), indicating evolutionary conservation of conspecific vocal recognition. In developing infants vocal recognition skills have been shown to precede language, and there is even evidence to suggest that neonates (Ockleford, Vince, Layton *et al.* 1988) and even preterm fetuses are able to distinguish familiar voices (Kisilevsky, Hains, Lee *et al.* 2003), quantified in heart rate measurements. Yet whereas studies of speech perception have enjoyed a great deal of neuroscientific interest, perception of paralinguistic information in voices has only attracted interest recently in cognitive psychology and neuroscience and far less is known about the neural bases.

Evidence from neuroimaging, neuropsychological and electrophysiological studies suggest that linguistic and paralinguistic vocal information are processed in partially dissociated functional pathways. Neuropsychological studies have been described in which voice recognition is impaired (phonagnosia) but comprehension of speech is intact (Van Lancker & Canter 1982; Van Lancker, Cummings, Kreiman *et al.* 1988), and the converse in which subjects have impaired comprehension of speech but demonstrate normal voice recognition (Van Lancker *et al.*1982). Double dissociations have also been reported in brain lesion subjects between the ability to recognise vocal emotions and comprehend speech: with reports of cases commonly with right hemisphere lesions demonstrating impaired emotion recognition but intact speech comprehension, and aphasic subjects with left-hemisphere lesions showing the opposite pattern (Barrett, Crucian, Raymer *et al.* 1999; Ross & Monnot 2008). Studies in healthy volunteers have also shown that speakers can be reliably recognised in voice samples where linguistic information has been eliminated (by temporal reversal or filtering the frequency content) but some or all of the spectrotemporal attributes are retained (Compton 1963; Pollack, Pickett, & Sumby 1954; Van Lancker, Kreiman, & Cummings 1985). Whereas speaker discrimination has been found to be

impaired when nonverbal spectrotemporal content is impoverished in whispered speech, where linguistic comprehension is retained (Pollack *et al.*1954).

Electrophysiological studies of healthy volunteers have shown selective early pre-attentive responses to nonverbal vocal stimuli when compared to either other complex sounds or verbal stimuli (Knosche, Lattner, Maess *et al.* 2002; Levy, Granot, & Bentin 2001), suggesting that there is early parallel processing of phonetic and paralinguistic vocal information. Specific electrophysiological responses have also been shown in response to speaker related vocal characteristics including the familiarity of the speaker (Beauchemin, De Beaumont, Vannasing *et al.* 2006), and the accent of the speaker (Scharinger, Monahan, & Idsardi 2011), and dissociable responses were demonstrated for changes in speaker and vocal affect (Toivonen & Rama 2009). These studies provide evidence that stable paralinguistic information about speakers is rapidly extracted from single words, a process that involves pre-attentive auditory processing and is hypothesized to depend on long-term representations of identity and other paralinguistic categories.

Neuroimaging studies in healthy volunteers have also provided evidence for neural specialization of non-linguistic voice processing in auditory cortex. Belin and colleagues found areas located on the upper bank of mid- STS bilaterally which respond more to voices than to other complex sounds (Belin, Zatorre, Lafaille *et al.* 2000; Fecteau, Armony, Joanette *et al.* 2004), whereas right lateralized anterior superior temporal activations have been found associated with tasks involving listening to familiar voices (Belin & Zatorre 2003; von Kriegstein & Giraud 2004). Voice-specific responses in temporal cortex have also been demonstrated in infants from 7 months of age (Beauchemin, Gonzalez-Frankenberger, Tremblay *et al.* 2011) suggesting that the neural mechanisms that underpin nonverbal voice processing emerge early in development. Studies in the macaque brain have also found comparable functional regions in the anterior superior temporal plane that respond preferentially to conspecific vocalizations and to the identity of conspecific individuals (Petkov, Kayser, Steudel *et al.* 2008). These studies provide evidence for the existence of voice-selective neural regions which may be the auditory analogue to areas of inferior temporal cortex in humans and monkeys which show preferential responses to faces over other visual stimuli e.g.(Kanwisher, McDermott, & Chun 1997).

1.2.2. Voice production: shaping of the vocal auditory signal

The paralinguistic features of the voice are determined by the mechanical characteristics of the source of the sound in the larynx, and the shaping or filtering that occurs both in the larynx and the vocal tract (see Figure 1.1). Phonetic information in English and many other non-tonal languages is primarily conveyed by formant frequencies produced by changing the conformation of the vocal machinery and paralinguistic features are also produced by static and dynamic filtering of the sound (see Figure 1.1). "Static" individual differences in speakers' voices occur as a result of variation in vocal tract anatomy (for example as a result of size differences between speakers), or cultural factors such as vocal habits in positioning of vocal filters (Karpf 2006), whereas "dynamic" changes of the vocal filters may occur for example as a result of affective experiences of an individual. Individual differences in perceptual vocal characteristics of pitch and timbre result from differing conformations of the vocal tract. Although the perceptual properties of sounds do not entirely map to the physical acoustic structure directly: pitch is primarily determined by the fundamental frequency of the sound, whereas the timbre or vocal quality is determined by the temporal pattern and strength of different frequencies in the sound above the fundamental frequency (f_0).

Figure 1.1. The human vocal tract.

Image removed for online publication of this Thesis

Air expelled from the lungs through the glottis creates a pressure drop, causing oscillation of the vocal folds of the larynx. The temporal periodicity of the waveform that is generated from this oscillation determines the lowest frequency of the voice or f0. The fundamental frequency determines the perceived pitch of a person's speaking voice and is dependent on the size (mass and length) of the vocal folds which are larger in men and adults than in females or children, resulting in lower f0 values. In addition to f0, the sound wave emitted from the larynx contains harmonic overtones which are multiples of the fundamental frequency, and subsequent filtering by the vocal tract enhances some frequencies and attenuates others producing peak frequencies also known as formant frequencies (Ghazanfar & Rendall 2008; Titze 2008). Other filters in the mouth including lips, tongue and jaw further shape and articulate the sound.

Figure modified from http://training.seer.cancer.gov/head-neck/anatomy

The vocal tract acts in a similar way to a musical instrument, shaping the auditory signal in a way which results in a highly harmonic sound, meaning that voices have peaks in energy at frequency values at regular time intervals, which can be seen on spectrograms (images of spectral information across time), for example in Figure 1.2. This distinguishes voices from many other sorts of natural sounds, such as the sound of rain or wind, or mechanical sounds such as engines or tools which have little harmonic structure, containing noise or energy across a wide range of frequencies (for example in Figure 1.2 below: the sound of a box being dropped). Analysis of the spectrotemporal content of sound (or timbre) is thought to be critical to the perception of voices (Belin *et al.*2004; Goll, Crutch, & Warren 2010; Warren *et al.*2006).

Figure 1.2. Spectrogram of a human voice (saying "My dad's tutor"), a box being dropped, and a flute playing a single note.

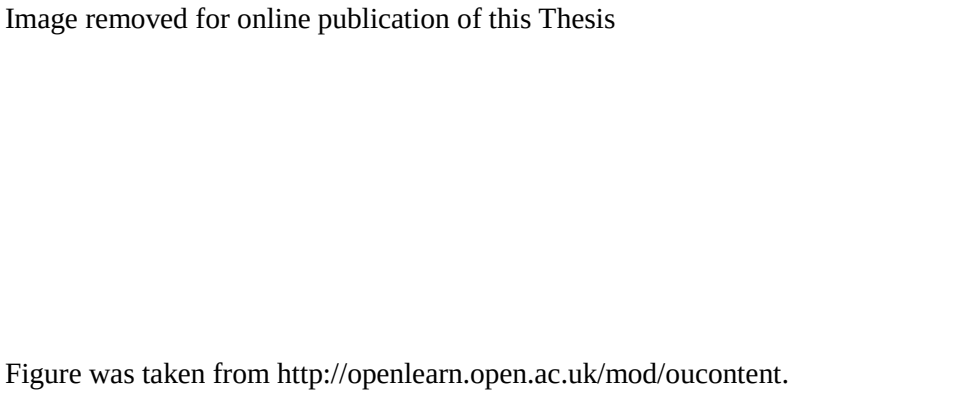


Figure was taken from http://openlearn.open.ac.uk/mod/oucontent.

1.2.3. Paralinguistic perceptual characteristics of voices

Perceptual properties of voices are likely to recruit distinguishable and separable, lower level cortical mechanisms to those that represent the object-level representation of the voice itself. The set of perceptual properties that bind together to encode a voice are underspecified. This is illustrated in comparison between the face inversion effect: in which presenting a face upside-down dramatically reduces its recognition, whereas in the voice modality it is not clear which are the critical features to "invert" or how to invert them (Bedard & Belin 2004). A contributing factor is likely to be variability in the vocal signal due to changes in the linguistic content, or changes in vocal quality as a result of social, cultural and other contextual factors such as mood

or physiological state (Cummings, Chin, & Pisoni 1996; Karpf 2006). Extracting meaningful paralinguistic information from such a variable stimulus is a complex problem for neurocognitive models of voice processing; this is exemplified in computerised systems for automatic speaker recognition which currently can only operate under highly constrained conditions. The rarity of case studies of perceptual voice processing deficits in the literature (Neuner & Schweinberger 2000; Van Lancker & Kreiman 1987; Van Lancker *et al.*1988; Van Lancker, Kreiman, & Cummings 1989) and a lack of detailed neuropsychological and auditory perceptual experimental analyses in those cases has meant that perceptual processing models for voices are much less well developed than for faces.

The contributions of perceptual cues to voice processing were investigated in early studies of healthy listeners' speaker, emotion or accent recognition ability, assessed after one or more perceptual features were removed or controlled by analysing or resynthesising the speech signal (Childers & Wu 1991; Compton 1963; Lavner, Gath, & Rosenhouse 2000; Pollack *et al.*1954; van Dommelen 1987). In other studies the acoustic features of the stimuli were analysed after healthy listeners categorised individual voices by speaker, emotion or accent; analysis typically involved visual inspection of vocal spectrograms, or by further listener ratings (Cummings *et al.*1996; Hanson 1997; Singh & Murry 1978; Walden, Montgomery, Gibeily *et al.* 1978). These studies tended to analyse a small closed set of voices limiting the generalisability of the results; however they have together provided a basis for understanding and testing voice perception, and have predicted differing but overlapping constellations of perceptual cues represent identity, accents and emotional content of voices (Juslin & Laukka 2001; Lavner *et al.*2000; Sauter, Calder, Eisner *et al.* 2010; Van Lancker *et al.*1985).

Based on this body of work, Belin and colleagues (Belin *et al.*2004) predict that a person's vocal identity is conveyed primarily through "static" or "invariant" perceptual features which characterize a speaker's unique vocal tract anatomy: primarily timbre, and also pitch cues including the modal range of pitches within which a person speaks (which highly overlap between speakers of the same gender), and formant frequency information embedded within or between formants. Emotion recognition is hypothesised to primarily depend on temporally varying or "dynamic" perceptual cues which include changes in intonation, intensity and duration, and to a lesser extent pitch and timbre cues. Recognition of accents (although not a feature of Belin's model) may utilize overlapping paralinguistic prosodic features to emotion recognition, such as prosodic cues (Berman *et al.*2003) (see Section 1.5 on accent processing), for

example segmental prosodic cues such as stress cues within a word, may be particularly important for accent processing.

1.3. Neuropsychological and neuroanatomical framework for assessing voice processing

Voices are rich in identity information, as described in Section 1.2.1. Nonverbal cues within a voice can tell us about the gender, size and age of the person, or more specific information about the identity or geographical background of the speaker. Gender, size and age discrimination are likely to depend on recognition of low level auditory perceptual cues which can be detected early in the auditory system (Baumann & Belin 2010; Childers *et al.*1991; Hanson 1997; Ives, Smith, & Patterson 2005; Klatt & Klatt 1990; Smith & Patterson 2005; Smith, Walters, & Patterson 2007; von Kriegstein, Warren, Ives *et al.* 2006). Tests of discrimination of gender and vocal size developed for this thesis are described in the Methods Section 2.4.1. Extraction of other information about the person such as their identity, accent, emotional state, intentions or mood from the voice, are likely to involve multiple levels of analysis involving perceptual, semantic, cross-modal and executive mechanisms. Recognition of familiar voices and accent processing are two aspects of voices investigated in this thesis that are likely to require perceptual analysis of multiple auditory features (such as prosodic cues for accent recognition and timbre and pitch analysis in speaker recognition; described in more detail in Section 1.2.3) and to involve mechanisms of semantic analysis. Cognitive, neuropsychological and neuroimaging studies of voice and accent processing will be considered in the following section, and a model for voice processing in this thesis is presented in Section 1.7.3.

1.4. Familiar voice processing: cognitive, neuropsychological and neuroanatomical mechanisms

1.4.1. Familiar voice recognition in healthy volunteers

A growing body of psychological studies have investigated familiar voice recognition: either using the voices of famous individuals (Hanley & Turner 2000; Schweinberger, Herholz, & Sommer 1997) or personally familiar people such as family members or colleagues (Nakamura *et al.*2001; Pollack *et al.*1954). Early studies of familiar voice recognition often used a small, closed set of familiar speakers and multiple-choice response formats, and it was found that recognition performance critically depends on the size of the response set (Pollack *et al.*1954; Saslove & Yarmey 1980). More recently, studies have used famous people as a way of assessing voice recognition in a large number of control subjects using many familiar identities, enabling an

"open" response procedure such as confrontation naming of individuals. Open response sets present a more realistic and ecologically valid test of recognition (Papcun, Kreiman, & Davis 1989) and parallel tests of familiar face recognition (Snowden, Thompson, & Neary 2004; Warrington & James 1967). Studies investigating the ability to judge familiarity (yes/no decisions), and semantic and episodic recall from voices (Damjanovic & Hanley 2007; Hanley, Smith, & Hadfield 1998; Hanley *et al.*2000; Schweinberger, Herholz, & Stief 1997), together they have demonstrated that a consistent and stable representation of a speaker can be reliably formed in normal individuals after very short exposures: even less than a second in duration (Schweinberger *et al.*1997), and that previously unheard voices can be learnt and remembered after a delay of several weeks (Papcun *et al.*1989).

Repetition priming studies in healthy volunteers have provided evidence that voice recognition may operate via similar cognitive mechanisms to faces (Ellis, Jones, & Mosdell 1997; Schweinberger 2001; Schweinberger *et al.*1997). In these studies, exposure to a sample of an individual's voice has been found to benefit familiarity decisions towards the same person's voice presented at a later time point, with increases in efficiency measured in terms of decreases in reaction time and error rates (Ellis *et al.*1997; Schweinberger *et al.*1997). Repetition priming effects have been shown to be resilient to relatively long time intervals (Ellis *et al.*1997) and to changes in the speech sample between presentations of the same individual (Schweinberger *et al.*1997), suggesting that familiar voice recognition involves activation of structural representations of familiar voices stores in long-term memory, which are insensitive to changes in speech content. Similar studies have found that familiarity decisions to a famous voice are also significantly faster if previously presented with the same person's face within a short time frame (Ellis *et al.*1997; Schweinberger *et al.*1997). These results suggest strong cross-modal connectivity between face and voice modalities, and provide evidence for a shared cross-modal processing stage where familiarity decisions are made across face, voice and name modalities (see Section 1.7.1 describing the Bruce and Young model of voice recognition).

Voice recognition in healthy controls has been generally found to be a slower and less successful process than face recognition (Damjanovic *et al.*2007; Ellis *et al.*1997; Schweinberger *et al.*1997). This is likely at least partially to reflect a lower frequency of exposure to famous voices in isolation compared with faces (Hanley *et al.*2000). Hanley and colleagues have also shown that controls are less able to recall person-specific information from voices they have judged as familiar, resulting in a large percentage of "familiar only" states from voices compared to faces

(Hanley *et al.*1998; Hanley *et al.*2000). Accordingly it has been proposed that there is weaker activation of associated semantic information from voices (Hanley *et al.*2000) as a result of weaker connections between the PIN (the gateway to accessing semantic information) and VRUs compared to FRUs (see Bruce and Young model of person recognition in Section 1.6.1). An alternative hypothesis is that familiarity mechanisms differ between auditory and visual modalities. The extent to which there is cognitive and neural segregation of voice familiarity and recognition from other modalities of person recognition is a contentious issue which will be investigated in this thesis, and is considered in relation to neuropsychological studies below (Section 1.4.2).

1.4.2. Disorders of familiar voice recognition: phonagnosia

Impairments of familiar voice processing have been described in studies of patients with focal temporal and parietal lesions (Ellis, Young, & Critchley 1989; Hanley, Young, & Pearson 1989; Lang, Kneidl, Hielscher-Fastabend *et al.* 2009; Neuner *et al.*2000; Van Lancker *et al.*1987; Van Lancker *et al.*1982; Van Lancker *et al.*1988; Van Lancker *et al.*1989) and more recently in developmental case KH (Garrido, Eisner, McGettigan *et al.* 2009) who reported a lifelong social problem in which she was unable to recognise the voices of family, friends and colleagues over the telephone. Phonagnosia is much less well characterised than prosopagnosia, and aside from the technical difficulty of assessing voice processing in clinical settings, this may be because phonagnosia is intrinsically less salient than face or name recognition deficits (Neuner *et al.*2000). Nevertheless, it may be a significant and disabling clinical issue, especially in situations where compensatory cues are reduced or unavailable (e.g., over the telephone).

Like deficits of familiar face recognition (prosopagnosia), apperceptive and associative agnosias have been described for voices. Apperceptive and associative processing stages are established concepts in the visual modality (e.g. (Warrington & James 1988)) and are also characteristic of models of auditory processing of other complex sounds (Goll *et al.*2010). Both apperceptive and associative mechanisms involve the formation of object level representations in which perceptual features are bound together into unified representations. Whereas apperceptive auditory agnosia refers to the inability to perceive or analyse a gestalt object representation prior to the attribution of meaning, associative auditory agnosia refers to the inability to associate an object representation with meaning (Goll *et al.*2010), auditory object agnosia is described in further detail in Section 1.6.2. Models of visual object processing hypothesize that perceptual representations are needed which provide descriptions of the way that the features of an object

and its three dimensional structure combine, which may involve non-linearity, emphasising some features and de-emphasizing others (Riddoch & Humphreys 1987; Riddoch & Humphreys 2003; Taylor & Warrington 1971; Warrington & James 1986; Warrington *et al.*1988). These structural representations are essential to "object constancy", enabling an object to be recognised across different contexts and viewpoints.

In parallel to visual object processing, models of auditory object processing (see Section 1.6. for a description of auditory objects) (Goll *et al.*2010; Griffiths, Kumar, Warren *et al.* 2007; Griffiths & Warren 2004) hypothesise that structural representations or auditory templates contain complex non-linear mappings between spectral and temporal components in which object-relevant features are emphasized; for voices this is likely to emphasise timbre (Belin *et al.*2004; Goll *et al.*2010; Warren *et al.*2006), whereas in the case of words for example this may emphasise fine-grain temporal analysis (Griffiths, Rees, & Green 1999; Jorgens, Biermann-Ruben, Kurz *et al.* 2008; Otsuki, Soma, Sato *et al.* 1998).

Vocal apperceptive deficits have been described in patients with focal lesions to either hemisphere that were unable to discriminate between two speakers of the same gender and/or perceive a voice under non-canonical listening conditions (Lang *et al.*2009; Van Lancker *et al.*1987; Van Lancker *et al.*1982; Van Lancker *et al.*1988; Van Lancker *et al.*1989). Deficits are primarily observed on both perceptual and semantic voice tasks in these studies (Neuner *et al.*2000; Van Lancker *et al.*1987; Van Lancker *et al.*1988), which can be explained by models which propose serial processing from perceptual to semantic analysis of the vocal signal (for example Bruce and Young's cognitive model described in Section 1.7.1). A small number of cases have also been described in which voice discrimination is impaired but familiar voice recognition is not (Van Lancker *et al.*1988), which is compatible with models developed for visual objects in which apperceptive processes can be recruited under some conditions (for example (Warrington *et al.*1988)). Double dissociations are needed in order to demonstrate functional independence of processing, and further detailed investigation of auditory perceptual deficits in such cases is required to further understand the auditory or cognitive mechanisms underpinning these deficits.

Cases of "associative" phonagnosia provide evidence that perceptual and post-perceptual semantic processes utilize functionally segregated neural mechanisms (for example Bruce and Young's cognitive model described in Section 1.7.1); those affected are unable to identify a

person from their voice or find them familiar (i.e. semantic mechanisms or associations are impaired), but performance on vocal perception tasks is spared (Neuner *et al.*2000; Van Lancker *et al.*1987). Associative phonagnosia has predominantly been described in cases with right-lateralized lesions affecting the temporal lobes often co-occurring with deficits in identification of familiar faces and names, in keeping with a loss of multimodal conceptual representations in the right temporal pole (Ellis *et al.*1989; Hanley *et al.*1989; Lang *et al.*2009). The function of the right ATL is debated; it may contain multimodal perceptual representations of people, amodal person-specific semantic concepts and/or autobiographical memory stores (Gainotti 2007a; Hanley *et al.*1989; Kopelman, Stanhope, & Kingsley 1999; Lucchelli & Spinnler 2008). Alternative theories propose that the right ATL represents "unique entities": individual known exemplars within a category (such as people, buildings or landmarks) that form separate nodes or "convergence zones" in which multiple simple perceptual features are bound together into complex multimodal configurations (Damasio 1990). Supporting this hypothesis, patients with focal lesions to the right ATL have also been described that exhibit impaired recognition of other visual unique entities: famous buildings and landmarks, although in general impairment on these tasks was less severe than at face recognition (Gainotti 2007b). It is unclear how the concept of "unique entities" extends to the auditory modality: the voices of familiar people are likely to contain hundreds of individual exemplars within a category, only people with specialised auditory expertise such as knowledge of musical compositions or bird calls may have comparable fine-grained knowledge in another category of sounds (Chartrand & Belin 2006; Chartrand, Peretz, & Belin 2008) .

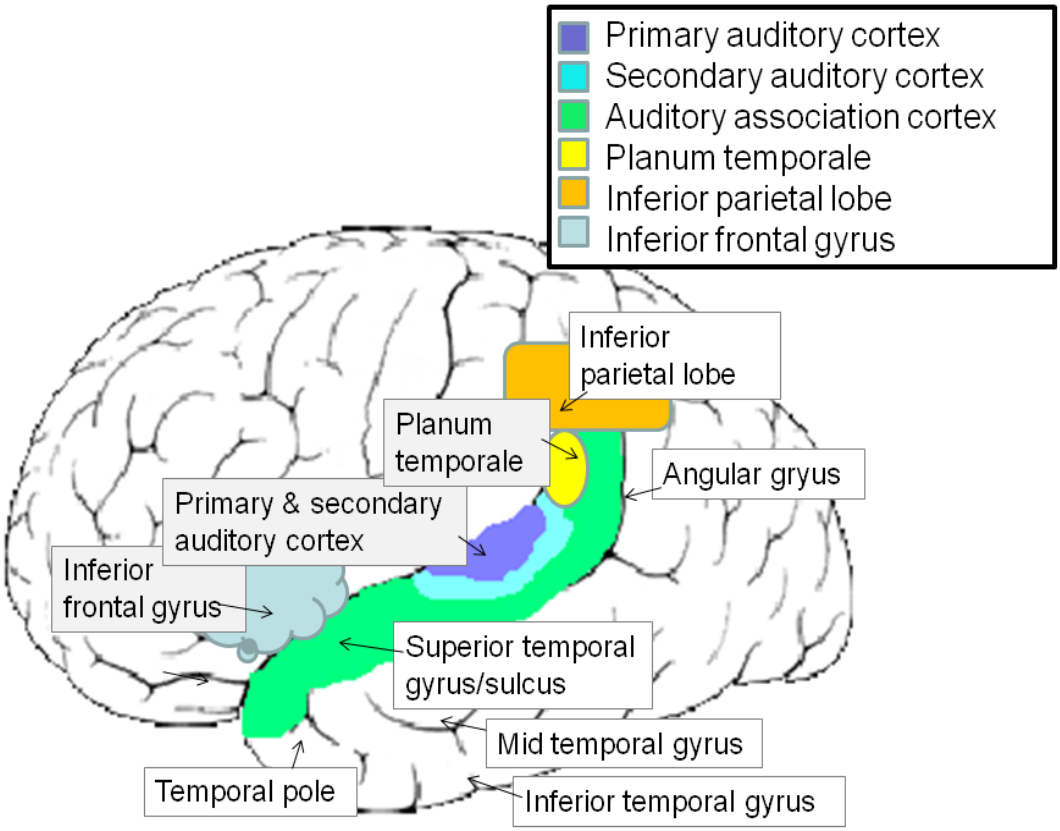
Phonagnosic cases have been described in which deficits have occurred independently of equivalent semantic deficits of person recognition in other modalities; most frequently phonagnosia with spared face recognition has been reported (Garrido *et al.*2009; Neuner *et al.*2000; Van Lancker *et al.*1982). These cases suggest there may be at least relative neural segregation between modalities at semantic levels of processing. Van Lancker and colleagues have attempted to look for more precise neural correlates of voice deficits and found deficits of familiar voice recognition correlated with damage in the right parietal cortex, whereas lesions to either temporal lobe were associated with speaker discrimination (Van Lancker *et al.*1988; Van Lancker *et al.*1989). The finding that unilateral temporal lesions to either hemisphere impair speaker discrimination is not inconsistent with functional imaging data which suggest bilateral or right-sided STS/STG involvement in perceptual analysis of voices (described further in Section 1.4.3). Right parietal cortex involvement in familiar voice recognition is more unusual, a region

which is generally associated with mechanisms of voice and auditory perception, such as memory for intonation, rather than semantic tasks (Berman *et al.*2003; Bishop & Miller 2009; Lattner, Meyer, & Friederici 2005; Rohrer, Sauter, Scott *et al.* 2010; Sokhi, Hunter, Wilkinson *et al.* 2005; Tucker, Watson, & Heilman 1977), and is not typically activated in functional imaging studies involving listening to familiar voices in controls (in Section 1.4.3). Methodological limitations, for example the lesion-deficit correlation methods which identified the presence or absence of damage by visual inspection of patients' CT scans was inaccurate or biased by selecting pre-specified regions of interest. Recent re-inspection of the CT scans in Van Lancker's study by Gainotti found that damage was associated with right inferior parietal or temporoparietal regions rather than other regions of the parietal lobe (Gainotti 2011), which is more consistent with data from neuroimaging studies (see Section 1.4.3). In this thesis, familiar voice processing will be assessed using multiple neuropsychological tasks with differing output demands, and VBM will be used to investigate the neural correlates of voice processing, a more powerful and less biased method of lesion-deficit analysis.

1.4.3. Functional imaging of voice recognition in healthy controls

An increasing number of studies have investigated voice processing in the healthy brain using functional imaging (Andics, McQueen, Petersson *et al.* 2010; Belin *et al.*2002; Belin *et al.*2000; von Kriegstein *et al.*2006) and have found a distributed network of areas engaged by voice processing tasks (Belin *et al.*2002; Imaizumi *et al.*1997; Nakamura *et al.*2001; von Kriegstein *et al.*2006), primarily recruiting regions of temporal lobe but also activations of inferior frontal and parietal cortices; see Figure 1.3.

Figure 1.3. Anatomical regions predicted to be involved voice processing on the basis of evidence from neuropsychological and functional imaging studies.



A number of studies have shown voice-specific activations when contrasted with other complex sounds. The first functional imaging studies of voice processing found voice-specific activations in bilateral mid STS in association with passive listening to vocal stimuli when contrasted with listening to complex stimuli that were spectrotemporally matched to voices, but not comparable in terms of their meaningfulness or saliency, such as musical bells and scrambled voices (Belin *et al.*2000; Binder, Frost, Hammeke *et al.* 2000). However since these first studies, voice specific activations in STS have been consistently shown using a range of contrast sounds including animal vocalizations and environmental sounds (Belin *et al.*2002; Fecteau *et al.*2004; Lewis, Talkington, Walker *et al.* 2009; von Kriegstein *et al.*2004).

Experimental designs manipulating the task rather than the acoustic material have examined the functional contributions of different neural regions during voice processing tasks. For example a study found that a non-linguistic task requiring subjects to judge if the voice is familiar to them or

not, was associated with activations in the right ATL when contrasted with a verbal task performed using the same spoken stimuli, (von Kriegstein, Eger, Kleinschmidt *et al.* 2003), providing evidence for specialised mechanisms of nonverbal vocal analysis in right anterior STS. It is unclear from several of these neuroimaging studies which of the voice processing stages implicated in cognitive models are associated with activations, as contrasts often involved comparison with acoustically but not semantically matched stimuli, (Belin *et al.*2003; Belin *et al.*2002; Belin *et al.*2000) or involved task manipulations which do not explicitly address these processing stages (von Kriegstein *et al.*2003; von Kriegstein *et al.*2004). This is likely to be at least in part due to the difficulty with finding acoustically and semantically valid control stimuli to compare to voices.

Hierarchical processing stages however have been predicted on the basis of task specific manipulations in fMRI paradigms. Perceptual mechanisms of voice analysis are implicated in mid and posterior regions of the STS by studies associating listening to unfamiliar voices when contrasted with other stimuli (Belin *et al.*2002; Belin *et al.*2000; Fecteau *et al.*2004). In one study a region in the right posterior superior temporal cortex activated to a greater extent to unfamiliar than familiar speakers, which it was hypothesised was due to a greater emphasis on spectrotemporal analysis of unfamiliar voices, (von Kriegstein *et al.*2004). These regions of activation overlapped with activations in response to other complex sounds temporally matched to speech, leading the authors to propose that this region plays a more general role in perceptual analysis of nonverbal temporally complex acoustic forms. In a different study, bilateral activations in posterior superior temporal cortices were found when subjects listened to a change in speaker, in a test comparable to the apperceptive tests used in neuropsychological studies (Warren *et al.*2006). In voice processing models regions in posterior STS are hypothesised to recruit generic auditory perceptual analysis mechanisms (Belin *et al.*2004; Scott, Rosen, Lang *et al.* 2006; Warren *et al.*2006).

More anterior regions in the upper bank of mid STS have been found to show voice-selective responses when contrasted with listening to other complex sounds (Belin *et al.*2002; Belin *et al.*2000; Fecteau *et al.*2004). Although the location of the voice-associated maxima in the STS varies between studies, and indeed between individual subjects (for example (Belin & Zatorre 2000)), these regions are hypothesised to contain voice-specific mechanisms of analysis and/or analysis of timbral characteristics that are highly relevant to voice processing (described further in Section 1.2.3). In Belin's model (Belin *et al.*2004) neurons here encode characteristic

configurations of voice-specific features, as implicated in nonhuman primates (Wang 2000), and correspond to the anatomical locus of the vocal apperceptive stage of processing. Functional connectivity analyses found that this region of the STS interacts with both posterior and anterior regions of the right STS (von Kriegstein *et al.*2004) and is likely to represent an intermediary stage of spectrotemporal analysis in the processing hierarchy.

Neuroimaging studies have also shown consistent voice-specific activations in response to familiar voices in right lateralized regions of the ATL, in keeping with the results of neuropsychological studies of associative phonagnosia (described in Section 1.4.2). Right ATL activations have been demonstrated when passive listening to familiar voices or performing familiarity judgements towards a set of known and unfamiliar voices (Andics *et al.*2010; Belin *et al.*2003; Imaizumi *et al.*1997; Nakamura *et al.*2001; von Kriegstein *et al.*2003; von Kriegstein *et al.*2004). Changes in right ATL activation have been associated with habituation or learning of previously unfamiliar voices, implicating a role for this region in forming long-term perceptual representations of speakers (Andics *et al.*2010; Belin *et al.*2003). As proposed by Belin (Belin *et al.*2004) it is plausible that the right anterior superior temporal lobe contains stored spectrotemporal representations of previously encountered voices.

The issue of whether there are modality specific representations of familiar people at the ATLs (see Section 1.4.2) has not been established using neuroimaging. A review of imaging of familiar voice and face processing found that voices recruit more superior regions and faces more inferior regions of the ATLs (Olson, Plotzker, & Ezzyat 2007) providing evidence that there is modality specific segregation. In contrast to neuropsychological studies which suggest that this region is involved in high level semantic representations (see Section 1.4.2), several imaging studies suggest that the right ATL may play a role in high level perceptual analysis of voices. Increasing the amounts of available spectrotemporal information in voice stimuli was associated with activation in the right ATL (Warren *et al.*2006), and in another study activations here were associated with familiarity discriminations towards both unfamiliar as well as familiar voices (von Kriegstein *et al.*2003; von Kriegstein *et al.*2004). Together this suggests the ATL plays a role in more fine-grained spectrotemporal analysis of voices, It is also possible that there is closer linkage between perceptual and semantic stages in the auditory modality compared to vision, as proposed in models of complex sound recognition (Goll *et al.*2010; Lewis *et al.*2009; Warren *et al.*2006).

Multimodal and cross-modal regions outside the temporal lobes have also been found to activate when presented with familiar voices (Imaizumi *et al.*1997; Nakamura *et al.*2001; Shah, Marshall, Zafiris *et al.* 2001; von Kriegstein *et al.*2003; von Kriegstein *et al.*2004) including the posterior cingulate (retrosplenial cortex) and precuneus, regions that are anatomically close to each other, and are unlikely to be voice specific; activating in one study also for familiar faces (Shah *et al.*2001). These regions are proposed to serve a role in multimodal integration, familiarity or imagery and are hypothesised to be recruited by temporal voice areas (Belin *et al.*2004).

Cross-modal responses in the fusiform face area and functional interactions between the STS and fusiform gyrus during familiar speaker recognition have been demonstrated in functional imaging paradigms (von Kriegstein *et al.*2006; von Kriegstein, Kleinschmidt, Sterzer *et al.* 2005). These results suggest that cross-modal coupling may occur at a perceptual level voice and face processing regions prior to semantic processing at the ATLs, and is proposed to be necessary for successful familiar speaker recognition. Studies in brain-damaged patients may help to clarify the role of inferior temporal cortex and other extra-temporal regions in familiar voice recognition.

Summary

Phonagnosia has been described as a developmental disorder and more commonly in association with focal damage involving the right or left temporal lobe or the right inferior parietal lobe. Associative voice recognition deficits have been described in the presence or absence of apperceptive deficits, supporting at least partially dissociable processing stages, which together with studies of repetition priming in healthy controls provide evidence for hierarchical mechanisms of voice recognition analogous to face processing models. Cognitive models of voice processing propose that low level perceptual analysis occurs prior to more complex analysis of individual voices (an apperceptive stage) which is followed by semantic associative processing. Limitations to the methods used in the few patient studies investigating the precise neural correlates of phonagnosia means that the critical neuroanatomical bases of perceptual and semantic voice processes have not been established. In particular, very few cases of apperceptive phonagnosia have been described, and functional imaging paradigms investigating voice perceptual mechanisms suggest that there are generic mechanisms for analysis of complex sounds in posterior regions of the superior temporal cortex, whereas more voice-specific mechanisms in bilateral mid STS may be recruited.

Associative phonagnosia has been consistently described in cases with focal and degenerative pathology affecting the right ATL. Although modality-specific deficits have been described in a few studies, deficits generally co-occurred with familiar face and name recognition impairments, implicating a role for this region in multimodal representations of familiar people. Neuroimaging studies of healthy controls also support a role for right lateralized anterior temporal regions in familiar voice processing, however a review of imaging studies suggests that there is segregation of modality specific representations in the anterior temporal cortices, with voice representations represented more superiorly and faces more inferiorly. Whether the right anterior superior temporal lobe is the neuroanatomical locus of multimodal semantic representations of familiar people or where high level perceptual representations are formed has not been established; imaging studies investigating the functional contributions of this region suggest it may be involved in perceptual as well as semantic processes. Other extra-temporal multimodal cortical regions have also been implicated in imaging studies of familiar voice processing, for example the precuneus and retrosplenial cortex. Addressing associative and apperceptive voice processes in patients with brain damage involving widespread cortical networks of regions, using VBM may help to clarify the roles of temporal and extra-temporal regions.

1.5. Non-native accent processing: cognitive, neuroanatomical and neuropsychological mechanisms

1.5.1. Non-native accent processing in healthy controls

Communicating with speakers with different accents is an important task that is performed routinely by the healthy brain. Accents signal important information about speakers, including geographical origins, ethnicity and social milieu. Extraction of this information requires analysis of segmental (phonetic and phonological) speech features (Clopper & Pisoni 2004a; Clopper & Pisoni 2004b; de Mareuil & Vieru-Dimulescu 2006; Evans & Iverson 2004; Floccia, Goslin, Girard *et al.* 2006; Howell, Barry, & Vinson 2006; Van Bezooijen & Gooskens 1999). Unlike familiar voice identification where linguistic cues are hypothesised to be secondary to extraction of paralinguistic features, studies in healthy volunteers have found that segmental phonemic cues, in particular vowel sounds, are critical to recognition of regional accents in American and English listeners (Clopper *et al.*2004a; Clopper *et al.*2004b; Howell *et al.*2006). Sharing similarities with recognition of emotions from voices, accent processing is also likely to involve analysis of prosodic features such as pitch contour, rhythm and stress patterns. A study in Dutch and English healthy volunteers found that listeners were still able to accurately recognise accents when only

prosodic features were retained, but found a greater reduction in accent recognition accuracy when phonemic information was removed (Van Bezooijen *et al.*1999). In another study Spanish and French listeners were more influenced by prosodic parameters (pitch contour and duration) than phonemic characteristics in detecting Spanish and Italian accents (de Mareuil *et al.*2006). It is likely that a configuration of linguistic and paralinguistic cues characterise accents: object-relevant features for accents may for example include formant frequency information (representing vowel sounds (Tanji, Suzuki, Okuda *et al.* 2003)) and intonational cues, as described above.

The extent to which linguistic and paralinguistic information are separately extracted from speech is not established (Nygaard & Pisoni 1998; Nygaard, Sommers, & Pisoni 1994; von Kriegstein, Smith, Patterson *et al.* 2010). Processing of accents is likely to be a computationally demanding, multi-component neural operation recruiting brain mechanisms separable from those encoding the verbal content of speech. As an aspect of human meta-linguistic communication, accent processing is likely to bear some similarities to processing of voice identity (Berman *et al.*2003; Clarke & Garrett 2004; Clopper *et al.*2004a; Remez, Fellowes, & Rubin 1997). Neuropsychological models of accent processing have not been developed. Like recognition of familiar voices (Belin *et al.*2004; Ellis *et al.*1997) recognition of accents is likely to involve perceptual analysis of the vocal signal and semantic mechanisms which associate vocal percepts with previously stored knowledge about geographical regions.

It has been found that healthy volunteers are able to recognise or categorize the region or country of origin of speakers based on a short speech sample with above chance accuracy in multiple choice arrays (Bayard, Gallois, Weatherall *et al.* 2001; Clopper *et al.*2004a; Van Bezooijen *et al.*1999), however performance has been shown to be very poor under free classification or open response procedures (Clopper & Pisoni 2007). Factors influencing the rates of recognition include the geographical boundaries chosen for the categories (e.g. countries or regions and areas within a country) and familiarity with an accent, whether through direct exposure to the accent (Clopper & Bradlow 2008; Clopper *et al.*2004a) or passive exposure to the accent through the media (Bayard *et al.*2001), suggesting that semantic representations of accents may be constrained by similar principles to learning other aspects of semantic knowledge, such as familiarity and frequency (Lambon Ralph, Graham, Ellis *et al.* 1998; Lambon Ralph, Graham, Patterson *et al.* 1999). Listener experience and exposure to a variety of accents (not necessarily the accents to be tested) improves accent discrimination performance (Clopper *et al.*2007), and it may be that listeners

develop skills at fine-grained vocal analysis, analogous to developing expertise in music listening or bird call discrimination (Chartrand *et al.*2008) in which listeners learn to discriminate and recognise relevant cues.

Psychophysical studies investigating the perceptual cost associated with comprehension of speech in the presence of an unfamiliar foreign or regional accent also find that familiarity with an accent reduces the magnitude of the processing cost associated with the non-native accent in the healthy brain (Adank *et al.*2009). Normal listeners show impressive flexibility in response to acoustic-phonetic confusions (Evans *et al.*2004; Norris, McQueen, & Cutler 2003) by shifting their phonetic boundaries for example to match the speaker's accent or idiosyncrasies (Evans *et al.*2004; Norris *et al.*2003). Processing costs measured either in terms of an increase in reaction times or error rates, have been observed particularly under adverse listening conditions such as speech presented in noise (Adank *et al.*2009; Best, McRoberts, & Goodell 2001; Clarke *et al.*2004; Floccia, Butler, Goslin *et al.* 2009; Floccia *et al.*2006).

In cognitive neuropsychological terms, a word or phoneme spoken in an unfamiliar (foreign or regional) accent has been viewed as an extreme form of native inter-speaker variation (Best *et al.*2001; Clarke *et al.*2004; Evans *et al.*2004; Floccia *et al.*2006; Nathan, Wells, & Donlan 1998; Schmale & Seidl 2009). Theories suggest that through exposure to multiple speakers of a native accent, and/or exposure to different accents tolerance regions develop around prototypes of phonetic elements (Floccia *et al.*2009; Francis, Nusbaum, & Fenn 2007; Goldstone 1994; Nathan *et al.*1998). The magnitude of the processing cost associated with a non-native accent is influenced by the accent's acoustical distance (e.g. phonological-phonotactic) from native speech (Clarke *et al.*2004): in which regional accents generally fall closer to native speech than foreign accents. Processing of foreign-accented speech therefore may be a noisier version of a pattern-matching process in which speech sounds are matched to smoothed spectral templates of vowels (for example (Hillenbrand & Houde 2003; Nearey 1997)). Processing of foreign-accented speech could be regarded as a 'non-canonical view' of a phoneme or word (or other auditory object) which will be accommodated by linguistic mechanisms for native speech with varying degrees of accuracy depending on an individual's previous experiences. A priori processing non-native accents may engage auditory apperceptive mechanisms analogous to the visual apperceptive mechanisms that process unusual views of visual objects (Goll *et al.*2010; Riddoch *et al.*2003; Warrington *et al.*1988) (apperceptive voice processes are described in Section 1.4.2).

1.5.2. Neuropsychology of accent processing

Recognition of regional or foreign accents has not to date been studied in brain-damaged subjects. Several studies have shown a reduced comprehension of unfamiliar or non-native accented speech in older adults (Adank & Janse 2010; Burda, Bradley-Potter, Dralle *et al.* 2009; Burda, Scherz, Hageman *et al.* 2003), in non-demented aphasic subjects (Burda, Brace, & Hosch 2007; Burda *et al.*2009; Dunton *et al.*2011) and in dementia patients (Burda, Hageman, Brousard *et al.* 2004). The majority of these studies failed to demonstrate a significantly greater cost relative to age-matched controls for the unfamiliar accent relative to the familiar (Burda *et al.*2004; Dunton *et al.*2011), this includes a study of patients with AD and vascular cognitive impairment (Burda *et al.*2004). The cognitive and neuroanatomical bases of deficits in these studies have also rarely been investigated. One study in aphasia speculated that impaired verbal comprehension or cognitive speed in the patient group were critical (Dunton *et al.*2011).

In any neuropsychological study of accent processing due consideration may be given to Foreign Accent syndrome (FAS), a rare form of speech apraxia typically described in stroke cases (Blumstein, Alexander, Ryalls *et al.* 1987; Hall, Anderson, Filley *et al.* 2003; Kurowski, Blumstein, & Alexander 1996) in which the patient produces speech sounds that are not part of the speaker's native language. It has most commonly been described in cases with left hemisphere lesions, but also in association with frontal, parietal, cortical and subcortical lesions. FAS has been described in a degenerative case of PNFA and associated with left perisylvian atrophy (Luzzi, Viticchi, Piccirilli *et al.* 2008). Although FAS is not a common symptom in PNFA, patients may share features of FAS such as dysprosody, distorted vowel sounds, and deviations in rhythm and stress.

It has been proposed that FAS is a listener-bound phenomenon rather than a syndrome (Kurowski *et al.*1996; Van Borsel, Janssens, & Santens 2005), based on detailed analyses of patients' errors which were all phonetically plausible sounds within their language (Kurowski *et al.*1996), variability in the accent listeners perceived the patient to have (Van Borsel *et al.*2005), and the common co-occurrence of other speech disorders (including aphasia, apraxia and dysarthria) which could account for features such as dysprosody and distorted speech sounds. It is notable that studies of FAS have not assessed subjects' own perception and comprehension of accents, therefore it remains possible that in some cases deficits may have been underpinned by deficits such as impaired feedback from the mechanisms used to perceive linguistic or paralinguistic cues.

1.5.3. Functional imaging of accent processing

Functional imaging evidence has implicated a distributed network including STS, STG, planum temporale, inferior parietal and inferior frontal gyrus in accent processing (Adank, Noordzij, & Hagoort 2012; Berman *et al.*2003). The components of this network are likely to mediate particular aspects of accent analysis, including vocal timbre (Belin *et al.*2000; Fecteau *et al.*2004), intonation (STG and STS) (Meyer, Alter, Friederici *et al.* 2002; Meyer, Steinhauer, Alter *et al.* 2004; Patterson, Uppenkamp, Johnsrude *et al.* 2002; Zhang, Shu, Zhou *et al.* 2010) and dynamic phonetic cues (left superior temporal lobe: (Buchsbaum, Hickok, & Humphries 2001; Chang, Rieger, Johnson *et al.* 2010; Jancke, Wustenberg, Scheich *et al.* 2002; Liebenthal, Binder, Spitzer *et al.* 2005; Scott *et al.*2006; Turkeltaub & Coslett 2010)). Whereas Berman's study (Berman *et al.*2003) found activations in a right lateralized network, whereas Adank and colleagues found that a change in accent correlated with regions of left STG (Adank *et al.*2012) typically recruited in speech processing. It is possible however that activity in the latter study was related to unnatural deviations or irregularities in the speech as the study used an artificial accent rather than natural accents. Accent processing will be investigated in this thesis using accented speech by native English speakers to look at perception of natural accent variation.

Accent related activations in Adank and colleagues' study (Adank *et al.*2012) also included areas implicated in neuroanatomical models of voice processing (Belin *et al.*2004) including right posterior, mid and anterior regions of STG. It is likely that accent processing mechanisms engage anterior temporal regions previously implicated in other dimensions of semantic processing, including recognition of voices (Belin *et al.*2003; Nakamura *et al.*2001; von Kriegstein *et al.*2004; Warren *et al.*2006), and might a priori align with other dimensions of person knowledge or with other kinds of geographically differentiated knowledge (Crutch & Warrington 2003; Crutch & Warrington 2010; della Rocchetta, Cipolotti, & Warrington 1998; Ellis *et al.*1989; Gainotti 2007a). In common with voice processing, accent recognition may involve formation of high level representations of vocal features associated with conceptual knowledge about countries or regions. Accent processing however does not require knowledge of individual identities and therefore may recruit separable cortical networks (as proposed in the model for this thesis in Section 1.7.3). Whereas familiar voice recognition has been found to recruit cortical regions implicated in multimodal processing (such as retrosplenial cortex and fusiform gyrus), accent processing may not have such strong neuroanatomical connections to representations in the visual modality.

In addition to activations in temporal (and also parietal) cortical regions which share commonalities with voice processing regions, imaging studies of accent processing implicate inferior frontal areas similar to those activated when healthy listeners process emotional and linguistic prosody (Buchanan, Lutz, Mirzazade *et al.* 2000; George, Parekh, Rosinsky *et al.* 1996; Mitchell, Elliott, Barry *et al.* 2003; Zatorre, Evans, Meyer *et al.* 1992). As detailed in Section 1.2.3 recognition of accents may require overlapping paralinguistic prosodic features to emotion recognition, such as pitch contour and rhythm and it is possible that analysis of these "dynamic" features in accent and emotion tasks places greater demands on auditory working memory than speaker recognition tasks which rely on "static" perceptual cues. It has been hypothesised for example that functional connectivity between the auditory cortex and inferior frontal cortex may be involved in the retrieval of auditory information in working memory (Demonet, Chollet, Ramsay *et al.* 1992; Zatorre *et al.*1992).

In models of emotional prosody recognition, functional connections between inferior frontal and inferior parietal lobe areas are important in storing intonational information. Although inferior parietal activations were not found in the two imaging studies of accent discrimination, connectivity with this region may be important in higher level judgements of accent recognition or comprehension of accented speech. A hierarchical neuroanatomical model has been theorized for recognition of emotional prosody in which following low level perceptual analysis, higher level perceptual analysis in mid superior temporal and right posterior STS represent meaningful suprasegmental prosodic sequences. This information is fed forward to bilateral inferior frontal cortex where cognitive evaluations and explicit emotional judgements are computed (Wildgruber, Ackermann, Kreifelts *et al.* 2006), potentially by focussing attention on behaviourally relevant auditory features (Schonwiesner, Novitski, Pakarinen *et al.* 2007) and involving interaction with working memory mechanisms in the inferior parietal cortex (Paulesu, Frith, & Frackowiak 1993; Wildgruber, Pihan, Ackermann *et al.* 2002; Wildgruber, Riecker, Hertrich *et al.* 2005). A similar hierarchical model to that hypothesised for recognition of emotional prosody may also apply to accent recognition and comprehension as hypothesised in the model of voice processing proposed for this thesis in Section 1.7.3. Understanding the contributions of different regions within the network implicated in functional imaging studies will benefit from the study of accent processing in brain-damaged subjects with degenerative pathologies affecting frontal and temporal cortical networks.

Summary

Accents represent a key interface between linguistic and paralinguistic processing, conveyed both in paralinguistic features of a voice and also "non-canonical views" of linguistic units in the realization of vowels and consonants. As an aspect of human meta-linguistic communication accent processing is likely to share similarities with processing of familiar voices. Neuropsychologically, recognition of accents may involve hierarchical perceptual and semantic mechanisms in which auditory percepts are associated with relevant geographical knowledge. Neuroanatomically, similar regions in the posterior and anterior superior temporal lobes have been implicated in functional imaging studies of voices and accent processing. Whereas studies of voice processing often show right lateralized patterns of activity in STS/STG, activations during accent processing are bilateral, including activations in left superior temporal regions implicated in linguistic feature analysis and speech intelligibility. Functional imaging studies of accent processing show commonalities with networks of activation found in analysis of emotional or linguistic prosody, in particular implicating inferior frontal regions which may reflect auditory working memory demands required for tracking intonational cues over time or comparing "non-canonical" phonetic cues in accent processing. Studies investigating comprehension of accented speech suggest that processing of unfamiliar accents may represent an extreme form of inter-speaker variability and may engage auditory apperceptive mechanisms, in particular access to high level perceptual representations of linguistic units. There are few neuropsychological studies of accent processing in the literature; investigation of the cognitive and neuroanatomical bases of accent processing deficits in brain-damaged subjects may enhance understanding of the role of particular regions within the networks identified in functional imaging studies.

1.6. Voices as auditory objects

In models of voice recognition, beyond basic perceptual parsing of the auditory stimulus voices are processed separately from processing of other complex sounds. As proposed earlier, this fits neatly with domain-specific hypotheses in the visual modality which predict dedicated neural and cognitive mechanisms for face perception and recognition (Kanwisher *et al.*1997). However, it has been shown that analysis of the spectrotemporal characteristics of voices shares commonalities with other meaningful categories of complex sounds (Singh & Theunissen 2003). Areas in the lateral and anterior temporal lobe contribute to processing of a variety of natural sounds including animal calls, environmental sounds and complex timbre processing (Lewis, Wightman, Brefczynski *et al.* 2004; Menon, Levitin, Smith *et al.* 2002; Thierry, Giraud, & Price 2003). In a functional imaging study in which contributions of vocal and non-vocal (musical

instrumental) perceptual information were varied using synthesised stimuli, behavioural ratings of the "naturalness" of stimuli rather than "voiceness" correlated with activations in mid-temporal voice areas (Belizaire, Fillion-Bilodeau, Chartrand *et al.* 2007). The extent to which voice perception and recognition mechanisms are separately processed compared to other complex sounds or "auditory objects" has not been established (Goll *et al.*2010; Leaver & Rauschecker 2010; Lewis *et al.*2009). In this section voice processing will be considered from the more general perspective of the processing of auditory objects, of which voices form one special category.

The concept of an auditory object itself is debated. One parsimonious definition is any meaningful pattern in the sound which can be disambiguated from a background auditory scene (Goll, Crutch, Loo *et al.* 2010; Griffiths *et al.*2004). Auditory objects include the acoustic source such as a voice or a musical instrument, or the acoustic events which emanate from the source which include speech content or melodies. Both the source and acoustic events present simultaneously (for example a vowel sound produced by an individual's voice), and both types of information need to be extracted by the auditory system. In order for an auditory object to be perceived and recognised, an invariant or shared configuration of features must be bound together and extracted to distinguish sound sources or events. Like voice processing models, models of auditory object processing predict a hierarchical organization. Early encoding of perceptual properties is followed by an "apperceptive" stage prior to the attribution of meaning, often tested by assessing the ability to discriminate between meaningful and meaningless complex sounds. An "associative stage" follows in which meanings or names are associated with perceptual representations, tested by the ability to identify different classes of sounds (Eustache, Lechevalier, Viader *et al.* 1990; Goll *et al.*2010). As is the case for voices, the neural mechanisms that represent stages of auditory object processing more generally are not established (Goll *et al.*2010; Griffiths *et al.*2007; Griffiths *et al.*2004).

1.6.1. Neuroanatomy of auditory object processing

Recognition of auditory objects is proposed to occur by extracting a set of 'acoustic signatures' or auditory features specific to a sound source through a hierarchically organized processing pathway along anteroventral auditory cortex. Analogous to ventral "what" (object-related), and dorsal "where" (spatial) pathways hypothesised in the visual system, distinct processing pathways have also been proposed in the auditory system (Kaas & Hackett 1999; Rauschecker 1998) on the basis of anatomical tract tracing, the functional properties of single neurons in the macaque, and

functional imaging studies in both nonhuman primates and humans. Although the functional organization of the human auditory system is debated (Belin *et al.*2000), there is evidence to suggest that auditory representations become increasingly complex from primary to secondary auditory cortex followed by association cortices: regions with distinct anatomical and functional properties, which include planum temporale, planum polare, STS/STG and insula. In primary auditory cortex, lesions cause deficits in sound detection (Heffner & Heffner 1986), and imaging in humans indicates there may be topographical representation of the intensity of sounds (Bilecen, Seifritz, Scheffler *et al.* 2002), and spatial mapping of frequency (tonotopy) of pure tones (Wessinger, Buonocore, Kussmaul *et al.* 1997). Sounds with increasing spectral and temporal complexity activate regions in nonprimary regions e.g. (Rauschecker 1998); secondary auditory cortex has been shown to activate in association with subjective pitch percepts and sequences (Griffiths, Buchel, Frackowiak *et al.* 1998; Schneider, Sluming, Roberts *et al.* 2005; Warren & Griffiths 2003; Warren, Uppenkamp, Patterson *et al.* 2003). The functional contribution of auditory association cortices to auditory object processing is unclear; it is possible that discrete regions represent specific features or combinations of features, perform specific analyses (responding to modulations of amplitude or frequency for example) or contain categorically organized auditory or semantic representations.

Essential to the concept of an auditory object is "object constancy" (described in Section 1.4.2 on apperceptive voice processes) in which invariant features or characteristics which define an auditory object are bound together to form a coherent percept, enabling an object to be detected in the presence of background noise and distinguished from other auditory objects when presented simultaneously with other sounds in the environment. In models of auditory object processing a specialised role has been proposed for a region in auditory association cortex in posterior STS: the planum temporale, which is proposed to act as a "computational hub", enabling analysis and segregation of spectrotemporal information from multiple sound sources as well as transformation between auditory and motor representations during speech (Griffiths & Warren 2002; Warren, Wise, & Warren 2005). Planum temporale is proposed to contain a generic mechanism for extracting spectrotemporal features for different classes of complex sounds, grouping the spectral and temporal components of a sound source into coherent percepts, enabling auditory object streams to be segregated and categorised or identified. It is hypothesized that the STS contains both generic as well as voice-specific mechanisms, such that vocal templates abstracted in more posterior regions are used in subsequent analyses more anteriorly in the STS where more detailed analysis of voice-specific perceptual and semantic information

occurs (Warren, Jennings, & Griffiths 2005; Warren *et al.*2006); such a hierarchy is consistent with models of voice recognition (Belin *et al.*2004; von Kriegstein *et al.*2004).

Analysis of timbre, such as spectral envelope or degree of periodicity (Lewis *et al.*2009) which are highly relevant to voice processing, is thought to contribute importantly to the spectrotemporal signature of individual auditory objects and has been shown to recruit several posterior auditory association areas as well as anterior STS/STG (Halpern, Zatorre, Bouffard *et al.* 2004; Warren *et al.*2005). It has been proposed that neurons encoding harmonic or spectrotemporal features in bilateral mid superior temporal cortex serve as an intermediate stage of processing between early analysis in Heschl's gyrus (in primary and secondary auditory cortices) and later processing in anterior temporal regions (Lewis *et al.*2009) where further analysis or grouping of features occurs leading to meaning or identification of sounds. It is as yet unclear what aspects of voice processing warrants specialised mechanisms in the anterior STS but these could include both fine-grained spectrotemporal analysis (Menon *et al.*2002; Warren *et al.*2005) and fine-grained semantic mechanisms: as familiar people are a densely individuated semantic category (Chartrand *et al.*2008). Cross-modal processing may also be important for speaker recognition (Gainotti, Ferraccioli, Quaranta *et al.* 2008; von Kriegstein *et al.*2005).

Other categories of auditory object may also show separable representations in superior temporal cortex. Category specific activation clusters have been shown not just for voices but also musical instrument sounds and the phonetic content of speech (Leaver *et al.*2010). Large acoustic differences between sound categories (i.e. differing perceptual demands of processing different categories) tend to confound studies attempting to isolate categorical auditory semantic mechanisms. Perceptual differences may be integral to differentiating auditory objects or categories of objects.

One model of auditory object processing proposes that perceptual and semantic mechanisms of analysis are highly interactive, involving both bottom-up and top down processes. According to the model, perceptual regularities (involving spectral and temporal feature analysis) drive auditory object categorisation whereas 'top down' mechanisms forge associations between acoustic properties (Goll *et al.*2010). A number of imaging studies have shown that separable distributed neural networks across low and high level auditory processing regions represent different auditory object categories such as living and nonliving sounds, melodies, animals and mechanical sounds (Engel, Frum, Puce *et al.* 2009; Giordano, McDonnell, & McAdams 2010;

Lewis, Brefczynski, Phinney *et al.* 2005; Saygin, Leech, & Dick 2010). For example selective responses to particular sounds (such as singers, cats and guitars) were found in bilateral primary and association auditory cortices (Staeren, Renvall, De *et al.* 2009) which were hypothesised to represent a configuration of simple and complex features of these auditory objects. These results are in line with property-based or embodied cognition models of semantic processing which propose that retrieval of a concept occurs through activations across distributed cortical networks representing perceptual features of a concept (as well as motor and affective components) (Martin 2007). Recognition of different auditory object categories is likely to recruit relevant processing regions: environmental sounds may be more dependent on generic spectral and temporal analysis mechanisms, whereas voices are highly dependent on higher level analysis of spectral shape (Goll *et al.*2010). Together this model may explain generic deficits extending across different categories of complex sounds as well as more specific auditory agnosias.

1.6.2. Disorders of voice recognition and auditory agnosias

There is a limited literature on category selective neuropsychological impairments within the auditory modality. Selective deficits for voice processing at associative levels of processing have been described both in lesion studies and a developmental phonagnosic case, in particular in the presence of spared recognition of environmental sounds (Assal, Aubert, & Buttet 1981; Garrido *et al.*2009; Neuner *et al.*2000; Peretz, Kolinsky, Tramo *et al.* 1994). Neuner and Schweinberger (Neuner *et al.*2000) described six cases with right and left hemisphere lesions with severe deficits of recognition of familiar voices (only one of whom also had a deficit at voice discrimination) that showed a preserved ability to recognise environmental sounds. The reverse dissociation: auditory agnosia for environmental sounds in the presence of preserved voice recognition has rarely been described, although two patients displaying this pattern were mentioned in a research report (Van Lancker *et al.*1982) these cases were not described in detail and require replication.

To date few studies have found dissociations between recognition of voices and other categories of sounds, for example there are no studies of phonagnosia in brain-damaged subjects reporting spared recognition of musical melodies or instruments, whereas conjoined deficits have been described in several temporal lobe lesion cases (Assal *et al.*1981; Peretz *et al.*1994). Developmental phonagnosic case KH was found to have intact recognition of both environmental sounds and familiar tunes (Garrido *et al.*2009): while it is possible that this reflects an abnormal pattern of auditory development, semantic or "associative" levels of auditory processing voice representations for familiar people may dissociate from semantic processing of other types of

auditory object. KH had no signs of brain damage on her MRI, and therefore the neuroanatomical correlates of selective associative phonagnosia are still to be determined.

Whereas there is some evidence for selective conceptual deficits in recognition of voices at an "associative level" there is little evidence for a selective perceptual deficit for vocal stimuli. An apperceptive case of phonagnosia was described that did not show impaired performance at environmental sound recognition (Neuner *et al.*2000), however further investigation of auditory perceptual deficits was not undertaken. Overlap has been found between vocal perceptual deficits and dystimbria in a patient with a right temporo-parietal lesion (Mazzucchi, Marchini, Budai *et al.* 1982). This case showed difficulty at both distinguishing and identifying voices, whereas perception of gender and recognition of other human sounds (including laughing and yawning) was unimpaired. Although an in-depth analysis of timbre processing was not performed, this subject showed deficits in recognising sounds from different categories requiring fine-grained distinctions of timbre: such as discriminating between human voices of the same age and gender or between vehicle engine sounds, but was not impaired at discriminating sounds distinguishable by rhythm, pitch or loudness cues. Other cases of dystimbria have been described in which perceptual phonagnosia is implicated; in a description of a stroke case with right posterior superior temporal lobe damage "human voices sounded 'unreal' as if they were being played through poor quality speakers" (Griffiths *et al.*2007).

Deficits of voice recognition have also been associated with impaired processing of melodic contour in cases with bilateral lesions to the superior temporal lobe (Assal *et al.*1981; Peretz *et al.*1994). These cases exhibited impairments of both recognition of musical melodies and vocal prosody but showed preserved performance on environmental sound and speech recognition tasks, hypothesised to be explained by spared rhythmic perceptual analysis (Peretz *et al.*1994). These studies, together with the rarity of selective apperceptive phonagnosia suggest that analysis of voices recruits perceptual mechanisms that are shared with processing of other auditory objects. In particular, voice processing is likely to be particularly reliant on aspects of timbre (in common with other auditory objects including nonhuman vocalizations and musical instruments), but also prosody or melodic contour (in common for example with musical melodies).

Summary

Whether deficits are selective for voices or extend to processing of other auditory objects is of interest both to voice processing models which treat voice processing as a separate module, and to

general models of auditory object processing. Such models implicate shared mechanisms for processing of different classes of auditory object, in particular a generic mechanism for extracting invariant perceptual signatures in posterior temporal cortex. Recognition of auditory objects is proposed to occur in an anteroventral "what" processing pathway in which increasingly complex auditory properties are extracted. Whereas some studies implicate category level processing of vocal and other auditory objects in STS/STG, others have identified distinct distributed networks of low and high level auditory processing regions for different kinds of objects, and propose that different levels of perceptual feature analysis are integrally recruited during recognition depending on the object or category of objects. Such a model may provide a basis for neuroanatomical differences between voice processing and processing of other auditory objects, in which voice processing is particularly highly reliant on analysis of timbre, which recruit mid temporal regions of the STS. This model is supported by neuropsychological evidence for dissociation between voice and environmental sounds at associative levels of processing but little evidence for selective apperceptive phonagnosia where deficits typically occur with other perceptual impairments, in particular those that require analysis of timbre. The neuroanatomical basis of associative auditory deficits that are selective for voices or extend to other sorts of sounds has not been established.

1.7. Models of voice processing

1.7.1. Bruce and Young cognitive model of person recognition

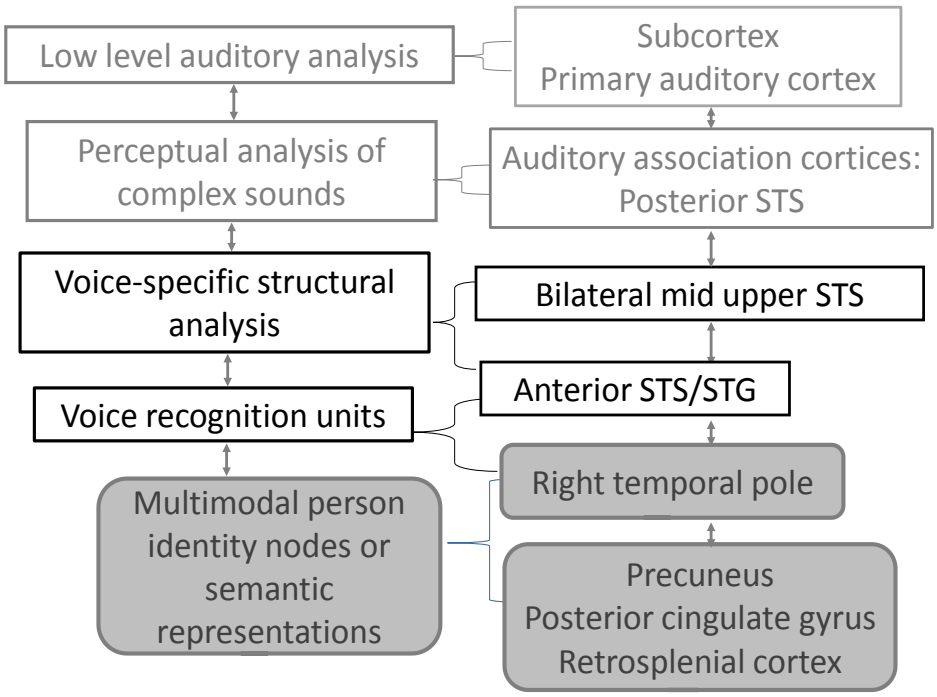
Cognitive models of voice processing have evolved from early models of person recognition (Bruce & Young 1986; Burton, Bruce, & Johnston 1990; Ellis *et al.*1997), and based on evidence from healthy volunteers and studies of phonagnosia these agree on the segregation of early unimodal perceptual processes (in which incoming face, voice and name stimuli are processed in parallel modality-segregated pathways) from later multimodal semantic mechanisms. The processing stages predicted in the model are displayed in Figure 1.4. Voice identification occurs via serially and hierarchically organised processing stages: basic auditory perceptual encoding of voices ("structural encoding"), is followed by further processing of feature information which is combined to form more complex structural descriptions of individual voices or templates called voice recognition units (VRUs). In parallel with face recognition units (FRUs), structural representations of familiar faces, VRUs enable a familiar voice to be recognised across listening conditions and linguistic contexts. Modality segregated perceptual representations from VRUs and FRUs are fed into cross-modal representations of familiar people or Person Identity Nodes

(PINs) which when activated enable subsequent retrieval of associated semantic information, such as the name or biographical information.

Whereas in the original versions of the model PINs were units that stored knowledge, in the updated "interactive activation and competition network" (IAC) model (Bruce *et al.*1986; Burton & Bruce 1993; Burton *et al.*1990) PINS do not store person specific information, but are cross modal nodes, serving only to signify a sense of familiarity in response to any modality of incoming stimulus. In this model, the PIN serves a crucial 'gating' function enabling access to biographical information which can be used to identify the person including the person's name. It acts as a modality-independent gateway: sufficient activation of the PIN from any modality facilitates access to semantic information units. The IAC model predicts that damage to the PIN will result in multimodal deficits such that familiarity and semantic judgements will occur equally in voice, face and name modalities. In the model, selective deficits of voice processing are only possible at the level of apperceptive processes at VRUs or earlier auditory perceptual processing. The occurrence of multimodal deficits of person knowledge in degenerative cases with damage to the right ATL (Ellis *et al.*1989; Gainotti, Barbier, & Marra 2003; Gainotti *et al.*2008; Gentileschi, Sperber, & Spinnler 2001) provides support for a neural region which plays a multimodal role in person recognition, either as the locus of the PIN in cognitive models or in higher level semantic representations.

Together the findings of neuropsychological and neuroimaging studies have provided evidence for the neural bases of processing stages predicted in the Bruce and Young model in which generic analysis of complex sounds occurs in posterior regions of the auditory association cortex, then structural encoding of voices occurs more anteriorly in bilateral mid STS, either involving voice-specific analysis or analysis of the harmonic structure of voices. Information is passed more anteriorly in the right anterior STS/STG where more fine-grained spectro-temporal analysis occurs and VRUs may be stored and extra-temporal multimodal processing regions are recruited. Figure 1.4 is an adapted version of a model proposed by Belin and colleagues (Belin *et al.*2004) hypothesized for voice recognition. The model is adapted to account for uncertainty of the roles of particular regions at higher levels of analysis in the right anterior STS/STG and temporal pole.

Figure 1.4. Anatomical regions predicted to be involved in cognitive processes of voice recognition



Model is adapted from the model proposed by Belin and colleagues (Belin *et al.*2004). Incoming voices are first processed at low levels of auditory analysis and the primary flow of information is downwards towards formation of a multimodal representation of a person at the last stages of analysis, although as indicated by the model connections are bidirectional. Grey boxes indicate generic auditory processing mechanisms, voice specific analysis is displayed in green, and multimodal mechanisms are indicated in blue.

1.7.2. Belin's model of voice processing

Elaborations of the Burton cognitive model has been applied to extend to other voice processing tasks (Belin *et al.*2004); hierarchical processing mechanisms (in which perceptual and semantic processes are segregated) are also applied to identification of vocal emotions and speech analysis, in line with other models (Scott *et al.*2006; Wildgruber *et al.*2006). Voice processing tasks are hypothesised to share basic auditory and spectrotemporal processing mechanisms, but operate in parallel at higher levels of perceptual analysis. The model therefore predicts that there is functional dissociation between speech analysis, recognition of emotions and identification of familiar individuals from their voice at "apperceptive" levels of processing. There are a number of cases described in the literature providing evidence for segregation of speech analysis from voice recognition processing and emotion recognition respectively, described in Section 1.2.1; however evidence for functional segregation of affective and identity pathways is more limited.

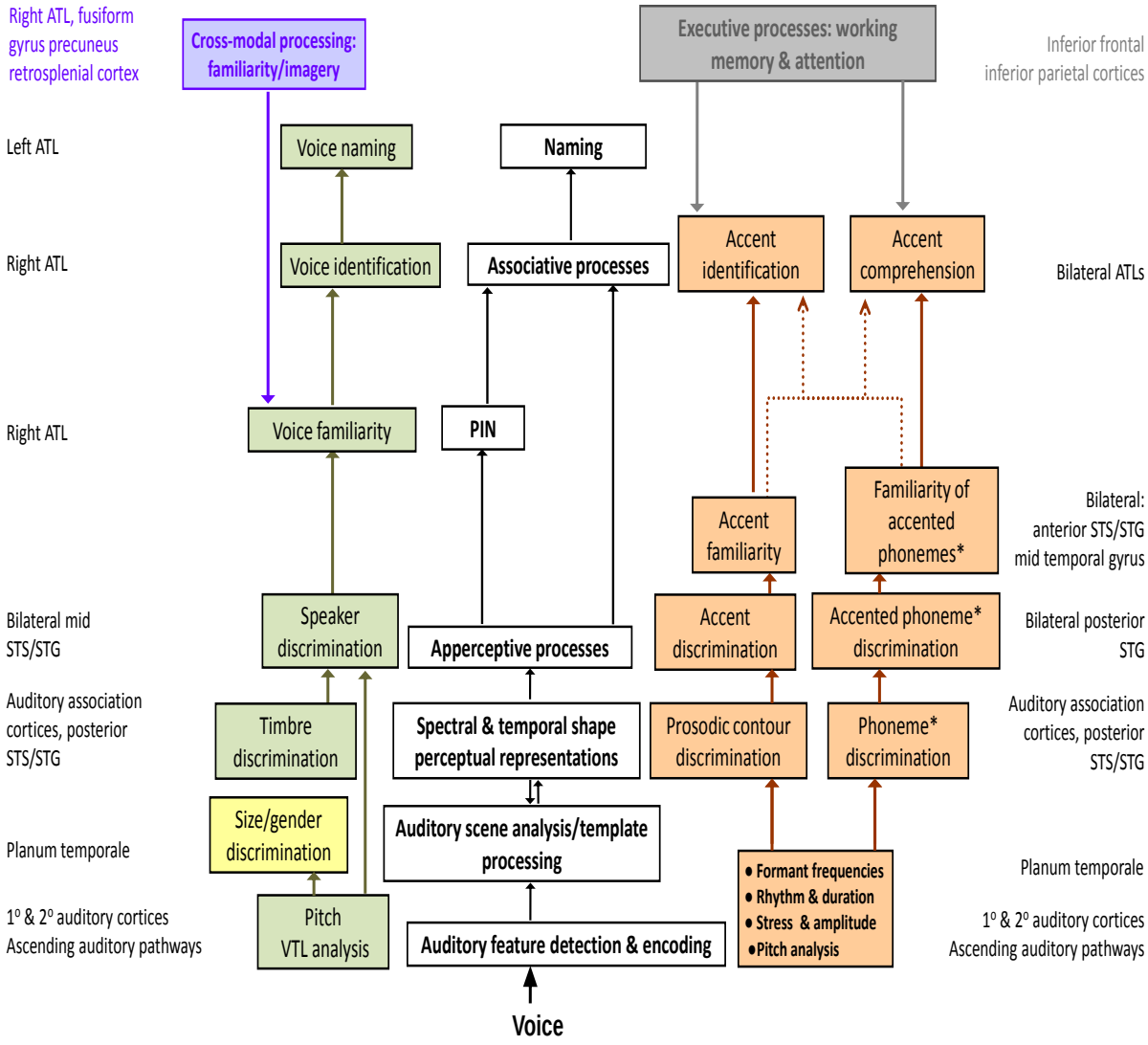
Developmental phonagnosic case KH (Garrido *et al.*2009) exhibited spared recognition of vocal emotions, and a less detailed case description of a patient with brain damage affecting the right temporal cortex indicates a similar pattern of results (Benzagmout, Chaoui, & Duffau 2008). No cases however have been reported that have shown the reverse dissociation: impaired affect or emotion judgements with spared identity processing.

Although face processing has offered a useful model for understanding recognition of familiar people from the voice, the extent to which these apply to other voice processing tasks has not been rigorously tested. Belin's model does not incorporate accent processing tasks and the model has been elaborated in this thesis to include accent recognition and comprehension, as described below in Section 1.7.3.

1.7.3. Model of voice processing in this Thesis

Together evidence from healthy controls and neuropsychological studies strongly suggest a hierarchical organisation in which auditory representations of voices and accents increase in complexity and are integrated with other cross-modal or executive cognitive processes. A simplified cognitive and neuroanatomical model is displayed in Figure 1.5 to provide a structure for understanding voice processing mechanisms underpinning accent processing and familiar voice recognition, based on Belin's model of voice processing and Goll and colleagues' neuropsychological model of auditory object processing (Belin *et al.*2004; Goll *et al.*2010).

Figure 1.5. Model of voice processing model proposed for this Thesis



This model was developed from previous models of voice and auditory processing (Belin *et al.*2004; Goll *et al.*2010). On the basis of neuropsychological and neuroanatomical evidence a hierarchical model of voice processing is proposed. Size and gender discrimination is indicated in yellow, and parallel processing pathways for voice identity processing (in green) and accent processing (in orange) are indicated with their component cognitive operations: utilising shared perceptual analysis at the bottom of the processing hierarchy, and at the top of the hierarchy separable mechanisms for each process, which are underpinned by differing neural substrates. Candidate anatomical substrates for these operations are displayed on the relevant side of the diagram, and although they are displayed as discrete "nodes", it is likely that areas cooperate as networks. Accented processing is proposed to recruit linguistic processing mechanisms, mechanisms are indicated by analysis of phonemes*, however the linguistic mechanisms may operate at the level of speech sounds: consonants and vowels, phonemes, phonological units and/or at the level of words: these processes were not specifically investigated in this thesis. Coloured arrows demonstrate the primary information transfer pathway and are uni-directional indicating the primary direction of communication between stages in the hierarchy, the dotted lines indicate that these perceptual cues may contribute to the processes that they are linked to, albeit to a lesser extent than those joined via solid arrows. Although the main pathways of information transfer are displayed, it is likely that connections may be reciprocal and that there is some lateral interaction between parallel pathways. Arrows linking template processing and spectral and temporal shape representations are shown as bidirectional to emphasise the dynamic updating of these templates via the interaction between incoming information and stored representations, in line with Goll and colleagues' model of auditory object processing (Goll *et al.*2010).

Mechanisms underpinning processing or recognition of accents do not feature in Belin's model but may operate via similar hierarchical mechanisms in parallel with a model of recognition of emotional prosody (Wildgruber *et al.*2006). In line with this model lower level auditory analysis would precede more complex accent specific perceptual representations and semantic and executive processes. On the basis of evidence from studies of accent processing in healthy controls (Section 1.5.1) it is hypothesised that comprehension of non-native speech would require the "speech analysis" processing pathway, in which the incoming speech must be matched to stored spectrotemporal templates of phonemic or phonological units of sound; the flexibility of these representations (if non-native speech represents a non-canonical representation of a phoneme as discussed in Section 1.5.1) would determine how successful the match is. Recognition of accents may also require reference to stored templates of spectrotemporal signatures of accents which contain a configuration of paralinguistic and linguistic cues represented in right and left ATLs. It is possible that accent comprehension is more reliant on left-lateralized mechanisms of linguistic analysis in posterior and anterior temporal cortices and inferior frontal cortices, whereas recognition of accents may be more dependent on nonverbal representations in the right ATL. In this model accent comprehension and accent recognition pathways have not been clearly disambiguated neuroanatomically due to limited neuropsychological evidence.

The model predicts that accent processes recruits a separate but overlapping neural network to voice processing. Shared neural mechanisms for voice and accent processes occur early on in the processing hierarchy: ascending auditory pathways and primary and secondary auditory cortices represent basic features which are combined into whole object representations in planum temporale. Bilateral mid and anterior temporal regions may be important in fine-grained analysis of timbre (such as spectral shape) which are critical to voice perception, whereas for regional and foreign accent analysis regions of bilateral mid and posterior temporal cortices are important in the extraction of meaningful linguistic segmental features and paralinguistic segmental and suprasegmental features, such as intonational, rhythm and stress cues. For both accent and voice processes, further auditory object analysis occurs in an anteroventral path along the STS and STG where more detailed perceptual and semantic information are extracted. The right anterior STS/STG may contain complex nonverbal spectrotemporal representations: templates of previously encountered voices (VRUs) and/or accents, or may contain semantic information about voices, such as knowledge about familiar people or geographical locations. More anteriorly, at the temporal pole multimodal perceptual representations of familiar people may be

formed (PINs) or amodal semantic knowledge about people or accents such as geographical knowledge could be stored.

In the last stages of processing, information is passed to extra-temporal regions which integrate high level auditory representations with other cognitive processes. In parallel with emotion recognition pathways, connectivity with inferior frontal and parietal regions may assist in tracking "dynamic" prosodic features of speech (including linguistic deviations or paralinguistic features) in working memory, or directing attention to behaviourally relevant perceptual features in processing of accents. In familiar voice processing, rather than depending on high level cognitive evaluations, other multimodal processing regions are likely to be recruited, such as the precuneus and retrosplenial cortex and the anterior fusiform. Whereas in Belin's model cross-modal processing mechanisms are emphasized for recognition of familiar people, recognition of regional and foreign English speaking accents are not likely to be so inherently tied to multimodal or facial information, and may be useful stimuli for isolating voice or auditory specific processes. Cross-modal connectivity may also occur in the voice recognition pathway earlier in the processing hierarchy between STS and face-processing regions in the inferior temporal cortex (von Kriegstein *et al.*2006; von Kriegstein *et al.*2005); these interactions are not displayed in the model.

1.8. Voice processing and neurodegenerative disease

The experiments designed in this thesis address the cognitive and neural mechanisms of voice processing in neurodegenerative disease. As described in Section 1.1 there are both clinical and neurobiological grounds for a systematic analysis of voice processing in cohorts of patients with degenerative dementia. Clinically, voice processing impairments are likely to be under-recognised but may contribute to socially disabling nonverbal symptoms. The break-down of voice processing in dementia and the concomitant atrophy within functional cerebral networks may enable identification of critical nodes within these networks and inform neurocognitive models and understanding of these diseases.

In this thesis voice processing will be investigated in four syndromes: SD, bvFTD, PNFA and typical AD. These syndromes are associated with distributed patterns of atrophy affecting temporal, parietal and frontal regions implicated in voice processing studies (Belin *et al.*2002; Imaizumi *et al.*1997; Nakamura *et al.*2001; von Kriegstein *et al.*2006; Warren *et al.*2006). Damage to distinct networks in these diseases results in unique patterns of cognitive impairment

and comparison of the cognitive and neuroanatomical bases of impaired voice processing between disease groups may facilitate understanding of the nonverbal symptoms that characterize each syndrome.

1.8.1. Voice processing in semantic dementia

SD is a subtype of FTLD and the paradigmatic disorder of conceptual knowledge which produces progressive disintegration of both verbal and nonverbal knowledge systems associated with relatively focal atrophy of the ATLs. Striking deficits of verbal semantic tasks (including naming, word fluency and comprehension) typically dominate the clinical syndrome, but impairments of nonverbal tasks are also described including impaired recognition of people (progressive prosopagnosia), agnosia for famous landmarks and places, and tactile and chemosensory agnosias (Edwards-Lee, Miller, Benson *et al.* 1997; Miller, Chang, Mena *et al.* 1993; Omar, Hailstone, Warren *et al.* 2010; Perry, Rosen, Kramer *et al.* 2001; Piwnica-Worms, Omar, Hailstone *et al.* 2010). Verbal and visual knowledge have been extensively investigated in SD, and evidence supports a multimodal breakdown of semantic processing. Early in the disease autobiographical memory is thought to be relatively spared (Neary, Snowden, Gustafson *et al.* 1998; Nestor, Fryer, & Hodges 2006) as well as other non-semantic abilities including visuospatial skills, working memory and arithmetic functions associated with relative sparing of parietal, occipital and dorsal frontal brain regions in this disorder (Edwards-Lee *et al.*1997; Thioux, Pillon, Samson *et al.* 1998; Waltz, Knowlton, Holyoak *et al.* 1999). The selective nature of the breakdown of conceptual knowledge in this syndrome offers an opportunity to investigate both the organization of person knowledge, and the organization of nonverbal auditory object concepts using voices.

ATL atrophy early in the disease is generally asymmetric and in the majority of cases presenting at specialist clinics and reported in the literature affects the left side more than the right (Brambati, Rankin, Narvid *et al.* 2009; Desgranges, Matuszewski, Piolino *et al.* 2007; Eustache *et al.*1990; Lambon Ralph, McClelland, Patterson *et al.* 2001). On MRI, atrophy early in the disease affects the anterior and inferior temporal lobe and spreads to the right ATL as the disease evolves. Cases where atrophy primarily affects the right anterior, inferior and superior temporal cortex: ('right temporal lobe variant' FTLD (right tvFTLD)) and subsequently affects the left temporal lobe suggest a more complex nosology. These cases typically display relatively spared language abilities initially, and a greater preponderance of nonverbal deficits; prominent clinical features are changes in behaviour and personality including obsessions, compulsions and loss of empathy (Brambati *et al.*2009; Busigny, Robaye, Dricot *et al.* 2009; Edwards-Lee *et al.*1997; Gorno-

Tempini, Rankin, Woolley *et al.* 2004; Perry *et al.*2001). Studies directly comparing performance in right and left –sided tvFTLD suggest that verbal and nonverbal semantic performance may dissociate: right sided ATL damage has been associated with greater deficits at face semantic tasks and left hemisphere ATL damage with greater impairments with names (Gainotti 2007a; Gainotti 2007b; Snowden *et al.*2004). Together differing anatomical and neuropsychological profiles justify using "tvFTLD" rather than SD to define patient groups. This view however is controversial, not least because both right and left temporal variant cases exhibit bilateral atrophy on magnetic resonance imaging (MRI) scans and display verbal and nonverbal symptoms as the disease evolves (Brambati *et al.*2009; Lambon Ralph *et al.*2001). The relations between verbal and nonverbal knowledge systems, and any segregation between left and right ATLs, are debated in theories of semantic processing (Humphreys & Riddoch 1988; Lambon Ralph & Patterson 2008; Snowden *et al.*2004; Warrington 1975) and have rarely been explored using nonverbal auditory stimuli.

The syndrome of progressive prosopagnosia is well recognised in association with right temporal lobe atrophy (Belin *et al.*2004; Chan, Anderson, Pijnenburg *et al.* 2009; Evans, Heggs, Antoun *et al.* 1995; Josephs, Whitwell, Vemuri *et al.* 2008; Joubert, Felician, Barbeau *et al.* 2003; Joubert, Felician, Barbeau *et al.* 2004). It is of considerable neuropsychological as well as clinical importance because it provides a window into the organisation of person knowledge in the brain (Bruce *et al.*1986; Burton *et al.*1993; Lucchelli *et al.*2008; Lyons, Kay, Hanley *et al.* 2006; Snowden *et al.*2004; Thompson, Graham, Williams *et al.* 2004; Warrington 1979) and is likely to represent a variant of SD dominated by deficits of nonverbal knowledge, including knowledge of familiar people (Gainotti 2007b; Gainotti *et al.*2008; Gentileschi, Sperber, & Spinnler 1999; Gentileschi *et al.*2001; Hanley *et al.*1989; Snowden *et al.*2004; Thompson *et al.*2004). Other channels of person knowledge, notably voices, commonly become affected with evolution of the progressive prosopagnosia syndrome (Gainotti *et al.*2003; Gainotti *et al.*2008; Gentileschi *et al.*2001), although voice recognition has only been rigorously tested in a few cases (Gainotti *et al.*2003; Gainotti *et al.*2008; Joubert, Felician, Barbeau *et al.* 2006). Voice processing is often anecdotally assumed to be normal in early progressive cases (Evans *et al.*1995; Gentileschi *et al.*1999; Joubert *et al.*2003), and is generally assessed only following the development of face recognition deficits (Gainotti *et al.*2003; Gentileschi *et al.*1999; Gentileschi *et al.*2001). Phonagnosia may not be identified as a clinical issue due to the lower saliency of voices but it may be that deficits in voice recognition co-occur with or precede difficulties in the face

modality. The modality specificity of deficits in dementia is relevant to predictions made by voice processing models (presented in Section 1.7.1).

Segregation of person-specific semantic representations, unique entities, or modality specific representations in the ATLs (described in Section 1.4.2 on phonagnosia) is disputed by theories proposing that bilateral ATLs are amodal stores or "hubs" with no category-specific organization. Evidence for this theory primarily comes from studies of visual and verbal object recognition in SD which show that category-specific deficits of semantic knowledge are very rare (Coccia, Bartolini, Luzzi *et al.* 2004; Humphreys *et al.*1988; Lambon Ralph *et al.*1999; Lambon Ralph *et al.*2008; Rogers, Ivanoiu, Patterson *et al.* 2006). Hubs contain abstract information about the similarity relations between semantic concepts which allow objects with very different perceptual features to be conceptually related. A number of cases of progressive prosopagnosia have been described in which there is relative preservation of other categories of knowledge, such as objects and animals (Evans *et al.*1995; Gentileschi *et al.*2001; Thompson *et al.*2004) and a case of left-tvFTLD showing the reverse pattern: spared person recognition in the presence of impaired general semantics (Thompson *et al.*2004). This evidence indicates that deficits of person knowledge can dissociate from other categories of semantic knowledge. According to some theories of semantic memory, category specific deficits can only occur as a result of underlying perceptual deficits (Humphreys *et al.*1988; Lambon Ralph *et al.*2008), however cases of selective prosopagnosia demonstrating spared perception of faces dispute this (Evans *et al.*1995; Gainotti *et al.*2003; Gainotti *et al.*2008; Gentileschi *et al.*1999; Gentileschi *et al.*2001). Perceptual voice processing has not previously been assessed in any degenerative cases. Performance on voice processing tasks represents a relatively unexplored avenue for investigating semantic and perceptual object representations in tvFTLD.

Impairments on other high level voice and auditory tasks have been described in SD including recognition of emotions from voices and from music (Omar, Henley, Bartlett *et al.* 2011; Rankin, Salazar, Gorno-Tempini *et al.* 2009), deficits of environmental sound recognition (Bozeat, Lambon Ralph, Patterson *et al.* 2000; Goll *et al.*2010) and recognition of musical instrument sounds (Omar *et al.*2010). It has been suggested that semantic deficits primarily underpin impairments of auditory object recognition, but apperceptive deficits may partially contribute at least to some auditory recognition tasks (Goll *et al.*2010; Goll *et al.*2010). Musical knowledge has been found to fractionate in SD, for example selective preservation of semantic memory for

musical melodies has been demonstrated with impaired knowledge of musical instruments (Hailstone, Omar, & Warren 2009; Omar *et al.*2010).

The core regions of atrophy in tvFTLD, the anterior and inferior temporal lobes, are areas implicated in voice processing both in lesion studies and functional imaging studies. As the disease progresses (both in cases with both greater left or right sided atrophy) it extends into the insula and orbitofrontal cortex: regions beyond the temporal lobe which have been implicated in voice processing studies in primates and functional imaging studies in humans (Fecteau, Belin, Joanette *et al.* 2007; Olson *et al.*2007; Remedios, Logothetis, & Kayser 2009; Wong, Parsons, Martinez *et al.* 2004). Investigating the neural basis of voice processing deficits in tvFTLD may improve understanding of the contributions of temporal cortices and extra-temporal regions to both voice and auditory object processing in SD as well as understanding of the symptoms these patients experience.

Summary

The selective nature of the breakdown of conceptual knowledge in right and left-sided tvFTLD offers an opportunity to investigate both the cognitive and neural organization of voice processing. Selective atrophy of the anterior and inferior temporal lobes in tvFTLD affects regions implicated in familiar voice recognition, and may extend to extra-temporal regions implicated in voice processing. Voice recognition deficits in tvFTLD however have only been described in the presence of multimodal deficits of recognition of familiar people in progressive prosopagnosia. The modality specificity of semantic voice processing deficits and the category specificity of any deficits within the auditory modality speak to the debate over the organization of semantic knowledge in the ATLs, for example the prediction that either person-specific knowledge or more general nonverbal semantic knowledge is represented in right ATL. Impairments of familiar voice recognition and recognition of other nonverbal auditory objects are assumed to occur at associative or semantic levels of processing: however perceptual processing of voices has not previously been tested in degenerative ases. It is possible that perceptual or apperceptive voice impairments (for example involving timbre processing) may contribute to voice recognition performance. Performance across perceptual and semantic voice tasks, and correlations with patterns of atrophy on MRI in tvFTLD may further understanding of the mechanisms of person recognition and nonverbal auditory semantic processes in this syndrome.

1.8.2. Voice processing in behavioural variant FTLD

BvFTD is a syndrome defined primarily by a decline in social function and personality which is typically associated with atrophy of the mesial and orbitofrontal cortex, anterior and medial temporal lobe structures, as well as limbic areas including the insula and amygdala (Piguet, Hornberger, Mioshi *et al.* 2011; Rascovsky, Hodges, Kipps *et al.* 2007; Seeley, Crawford, Rascovsky *et al.* 2008). A decline in emotional responsivity and interpersonal skills such as social conduct, social disinhibition and loss of empathy are some of the behavioural changes that appear early in the course of the disease (Neary *et al.*1998; Neary, Snowden, & Mann 2000). Other core behavioural features include stereotyped behaviours and alterations in eating patterns. Neuropsychological performance deficits may be relatively restricted to frontal executive and social cognition tasks in the early stages, and in particular episodic memory may be spared. BvFTD however is clinically heterogeneous, varying for example in the extent and severity of language impairment that appears with disease progression or the patterns of behaviours exhibited (for example apathy versus disinhibition (Snowden, Bathgate, Varma *et al.* 2001; Wicklund, Johnson, Rademaker *et al.* 2007)). Variation in behavioural presentation may mirror anatomic variability in the pattern of fronto-temporal-limbic atrophy (Whitwell, Przybelski, Weigand *et al.* 2009).

Regions implicated in functional imaging and lesion studies of voice processing are affected in bvFTD, including ATLs, amygdala, and also temporo-parietal atrophy (Seeley *et al.*2008; Whitwell *et al.*2009). Fronto-temporal networks that are affected in this syndrome overlap with those implicated in recognition of emotional and linguistic prosody and accent processing in functional imaging studies of healthy controls (Adank *et al.*2012; Berman *et al.*2003; Mitchell *et al.*2003). Impairments are typically reported on complex multimodal tasks involving detection of social faux pas or sarcasm for example (Kipps, Nestor, Acosta-Cabronero *et al.* 2009). Few studies have explored voice processing in isolation. Impairments of recognition of basic vocal and facial emotions however have been described (Keane, Calder, Hodges *et al.* 2002; Omar *et al.*2011; Snowden, Austin, Sembi *et al.* 2008), which in combination with impaired recognition of familiar people may contribute to the early emotional and social behavioural features and a lack of social connectedness characteristic of the syndrome (Chan *et al.*2009; Olson *et al.*2007; Omar, Rohrer, Hailstone *et al.* 2010; Rosen, Perry, Murphy *et al.* 2002; Rosen, Wilson, Schauer *et al.* 2006; Snowden *et al.*2008).

Nonverbal symptoms in bvFTD have been associated with atrophy of the right temporal lobe (Chan *et al.*2009; Edwards-Lee *et al.*1997; Perry *et al.*2001). Fronto-temporal atrophy is typically bilateral and symmetric; however asymmetric right-sided atrophy has also been reported (Seeley *et al.*2008) showing pathological and behavioural overlap with SD. Both syndromes are associated with socio-emotional dysfunction (Neary *et al.*1998; Neary *et al.*2000; Snowden *et al.*2001): whereas symptoms in SD have been linked to a fronto-temporal network mediating conceptual knowledge about objects, bvFTD has been associated with disruption to an anterior cortical network involving the temporal pole, orbitofrontal cortex and amygdala (Kipps *et al.*2009; Olson *et al.*2007; Omar *et al.*2011; Rosen *et al.*2002; Seeley *et al.*2009; Seeley, Menon, Schatzberg *et al.* 2007; Zhou, Greicius, Gennatas *et al.* 2010), which mediate evaluation of and responses to emotional stimuli and complex real-life social contexts (Kipps *et al.*2009; Olson *et al.*2007; Omar *et al.*2011; Rosen *et al.*2002).

Investigation of voice processing offers an opportunity to assess both object and emotion processing using a highly socially salient stimulus. Correlation between anatomical and behavioural performance in this thesis will offer an avenue to assess relative contributions of regions within the fronto-temporal-limbic network. Increased understanding of the breakdown of auditory cortical functions may facilitate an understanding of similarities or differences between SD and bvFTD syndromes, and in particular some of the unusual disorders observed. For example, it has been proposed that altered emotional reactions to music and sounds (Boeve & Geda 2001) may result from abnormal coupling between the auditory object property of timbre and affective processing mechanisms (Hailstone, Omar, Henley *et al.* 2009).

Summary

BvFTD presents primarily with social, emotional and behavioural changes, and it is possible that deficits of voice processing contribute to the dysfunctional social behaviours observed. Although voice processing has rarely been tested in bvFTD, atrophy involves a network of frontotemporal regions that are implicated in both familiar voice recognition, recognition of emotional and linguistic prosody and accent processing. The syndrome shows overlapping nonverbal symptomatology to tvFTLD related to involvement of the right ATL. Whereas in tvFTLD deficits have been associated with impaired object and person recognition, in bvFTD altered responses and representations of social and emotional stimuli have been associated with disruption to a fronto-temporal-limbic network. Voice processing offers an opportunity to analyse the

component deficits that may contribute to abnormal social behaviours in bvFTD using a highly socially salient stimulus.

1.8.3. Voice processing in Progressive Non-Fluent Aphasia

PNFA is a rare syndrome of FTLD typically associated with relatively focal atrophy around the sylvian fissure, particularly affecting the left inferior and dorsolateral prefrontal regions and insular cortex, and extending into STS (anteriorly and posteriorly) and inferior parietal areas (Grossman, Mickanin, Onishi *et al.* 1996; Mesulam 2001; Rohrer & Schott 2011). Degeneration of the perisylvian region has been critically associated with the striking language-led clinical features of the syndrome: effortful and dysfluent speech production, agrammatism and anomia, and research in PNFA has primarily focussed on the production and processing of verbal material. Quantitative observations of speech production errors in PNFA suggest that speech dysfluency is not necessarily caused by speech apraxia or motor planning impairments (Gorno-Tempini, Dronkers, Rankin *et al.* 2004), but may result from impaired syntactical processing due to agrammatism or working memory impairments (Gunawardena, Ash, McMillan *et al.* 2010; Wilson, Dronkers, Ogar *et al.* 2010), or due to impoverished phonological representations of words, resulting in phonemic paraphasias (Ash, McMillan, Gunawardena *et al.* 2010; Croot, Patterson, & Hodges 1998). Heterogeneity in the pattern of symptoms and foci of perisylvian atrophy is a feature of the disease (Gorno-Tempini *et al.*2004; Rohrer *et al.*2011). Despite very severe impairments of speech production, relatively spared single word comprehension, object recognition, and semantic task performance at least early in the course of the disease, differentiates the syndrome from the other language-led variant of FTLD, SD (Grossman & Ash 2004; Neary *et al.*1998).

An uncertain proportion of PNFA patients present with symptoms of auditory dysfunction. Cases with PNFA-like syndromes have been described in which either a selective deficit for the perception of words (word deafness) or agnosia for sounds and words has led the clinical presentation. Word deafness is rarely reported (Jorgens *et al.*2008; Otsuki *et al.*1998) and is not a clinical feature of all PNFA cases but has been associated with atrophy affecting posterior superior temporal lobes (either bilaterally or in the left hemisphere), regions implicated more generally in auditory object processing. Reports of category-specific auditory deficits affecting recognition of environmental sounds (Uttner, Mottaghy, Schreiber *et al.* 2006) and receptive linguistic and affective prosody have been described in PNFA (Rohrer *et al.*2010). Although

potentially less clinically salient, deficits of 'metalinguistic' functions including nonverbal processing of voices, may be particularly debilitating in patients with language impairments.

A limited number of studies have addressed apperceptive and associative levels of nonverbal auditory processing deficits (Goll *et al.*2010; Otsuki *et al.*1998; Rohrer *et al.*2010) and have implicated both perceptual and semantic impairments in PNFA. Studies have found that perceptual deficits can contribute to impairments at recognition of environmental sounds (Goll *et al.*2010), and impairments of linguistic prosody in this syndrome (Rohrer *et al.*2010). In a single case of PNFA with word deafness, concurrent impairments were observed on a perceptual task (temporal auditory discrimination), environmental sound recognition and syllable discrimination (Otsuki *et al.*1998). In this case it is likely that temporal perceptual deficits underpinned verbal and nonverbal sound recognition deficits, which is plausible in view of hierarchical models of auditory object processing (described in Section 1.6.1).

The neuroanatomical basis of auditory processing deficits in general has been inferred on the basis of the pathological disease process in PNFA rather than through lesion-behaviour correlations. Both word deafness and nonverbal sound deficits have been associated with posterior and/or anterior superior temporal lobe atrophy (either bilaterally or in the left hemisphere), regions implicated in more general roles in auditory processing. In a group study of PNFA, apperceptive environmental sound deficits was hypothesised to relate to the extent of posterior perisylvian atrophy (Goll *et al.*2010). In particular damage to planum temporale, which is predicted to be involved in either the formation or access to spectrotemporal templates of auditory objects (including words and voices) would provide a neural mechanism for deficits of receptive linguistic and paralinguistic auditory processing deficits as well as speech production deficits, as this region is implicated in the transformation between auditory and motor representations during speech (Warren *et al.*2005).

Lesion-behaviour correlation studies in PNFA are uncommon, which is likely to be at least in part due to the rarity of the syndrome. In the only VBM study of nonverbal voice processing in PNFA to date, a network of perisylvian regions involving frontal, temporal and parietal cortices was associated with performance on affective and linguistic prosody tasks across PPA syndromes (Rohrer *et al.*2010). These networks have been implicated in functional imaging studies of both emotion processing and accent processing (Adank *et al.*2012; Berman *et al.*2003; Mitchell *et al.*2003; Wildgruber *et al.*2006). Exploring voice processing in this group will increase

understanding of receptive auditory processing deficits in PNFA, and may help to delineate the functional contributions of perisylvian regions.

Summary

In PNFA deficits of language production dominate the clinical presentation; however receptive nonverbal auditory processing deficits have also been described which may contribute to social disability in the condition. Perisylvian atrophy in the disease affects posterior temporal regions which are critical to high level analysis of spectrotemporal information, and may underpin deficits in perception of paralinguistic and linguistic characteristics of voices. Receptive deficits for recognition of words and processing of nonverbal auditory objects (such as environmental sounds and emotional prosody) have been described in PNFA and group studies indicate that apperceptive and semantic deficits may occur. Such studies have rarely investigated the neural correlates of voice processing and in the only study to date, a network of frontal, parietal and temporal regions was implicated. Improved understanding of the neural regions involved in voice processing may help to delineate the contributions of different perisylvian regions to the relatively heterogeneous clinical profile of PNFA.

1.8.4. Voice processing in typical Alzheimer's disease

AD is clinically, neuroanatomically and neuropsychologically distinct from FTLD syndromes, and typically presents in a more clinically and pathologically homogeneous way. Commonly the first and most salient symptoms are memory dysfunction and pathological changes in the medial temporal lobes, in particular in the entorhinal cortex and hippocampus (Barnes, Ourselin, & Fox 2009; Lee, Buckley, Gaffan *et al.* 2006). The medial temporal lobes are thought to play a crucial role in the acquisition of long-term memories (Squire, Stark, & Clark 2004), and atrophy here is thought to result in pervasive deficits in episodic memory in AD. In some patients episodic memory function may present as an isolated symptom for many years; more usually however other cognitive functions including language, visuospatial skills, executive function and praxis, are affected either at presentation or with the involvement of other cortical regions (including parietal and frontal lobes) as the disease progresses.

The uniformity of AD is often emphasized, however Snowden and colleagues have found heterogeneity in the clustering of symptoms that present together (Stopford, Snowden, Thompson *et al.* 2008). For example aspects of memory loss may dissociate: deficits on tests of episodic memory (such as recall and recognition memory tasks), working memory tasks (involving the

retention of material across very short delays) or tests of semantic memory (assessed in tests of word or sentence comprehension, category fluency or naming) may not cluster together. A more dramatic illustration of this heterogeneity is the existence of posterior and language-led AD variants (Gorno-Tempini, Brambati, Ginex *et al.* 2008; Lehmann, Barnes, Ridgway *et al.* 2011; Lehmann, Crutch, Ridgway *et al.* 2009; Rohrer, Rossor, & Warren 2012).

Recognition of famous faces has been frequently assessed in AD, and in combination with deficits of familiar name recognition suggests that semantic deficits occur across modalities of person knowledge (Greene & Hodges 1996). Deficits have also been described on other visual and verbal semantic tests (Hodges, Salmon, & Butters 1992; Lambon Ralph, Patterson, Graham *et al.* 2003; Perry & Hodges 2000), and it is not clear whether these deficits are the result of generalized damage to conceptual representations as has been predicted in SD as a result of atrophy in ATLs, or due to impaired access to knowledge stores (Hodges *et al.*1992; Reilly, Peelle, Antonucci *et al.* 2011), which may for example relate to damage to temporo-parietal cortices (Noonan, Jefferies, Corbett *et al.* 2010). One reason for the lack of clarity is that semantic memory is often assessed using tests of object naming, and confrontational naming impairment in AD may have a different cognitive and neural basis to semantic deficits, for example due to impaired lexical retrieval as a result of temporo-parietal abnormalities (Stopford *et al.*2008). The contribution of perceptual deficits to person and object recognition deficits has seldom been studied in AD. The pattern of deficits on perceptual and semantic tasks and the neuroanatomical correlates of performance may help to elucidate this; for example the ATLs have been predicted to be the locus of amodal conceptual representations whereas temporo-parietal regions may either implicate underlying auditory perceptual deficits or semantic access impairments.

Recognition of voices has not been previously tested in AD, however impairments of recognition of other auditory objects, including words (Eustache, Lambert, Cassier *et al.* 1995), environmental sounds (Rapcsak, Kentros, & Rubens 1989) and musical melodies (Baird & Samson 2009; Omar *et al.*2010; Vanstone & Cuddy 2010) have been described, while recognition of musical emotions has been found to be spared. Reports of deficits of auditory processing in AD are relatively infrequent in the literature. A limited number of studies suggest that impairments arise at auditory perceptual levels, for example impairments of timbre discrimination (Kurylo, Corkin, Allard *et al.* 1993) or representation of auditory duration (Hellstrom & Almkvist 1997) have been described, although results have not be consistently replicated.

A recent study (Goll, Kim, Ridgway *et al.* 2012) suggests that AD impacts on more complex auditory processes in particular related to analysis of auditory scenes (in perceptual grouping tasks), impairments that were associated using VBM with atrophy in the posterior cingulate and posterior superior temporal lobe, the latter is in accordance with the proposed role for the planum temporale in auditory stream segregation (see Section 1.6.1). Other high level voice processing impairments have also been reported, for example deficits in discriminating linguistic and vocal emotional prosody (Allender & Kaszniak 1989; Roberts, Ingram, Lamar *et al.* 1996; Taler, Baum, Chertkow *et al.* 2008; Testa, Beatty, Gleason *et al.* 2001), which have been shown to present early in the disease and become more severe with disease progression (Testa *et al.*2001). Vocal impairments potentially present prior to deficits in language comprehension: one study proposed that linguistic meaning and context was used to compensate for impairments in interpreting prosodic cues in mild AD (Taler *et al.*2008). The cognitive and neuroanatomical underpinnings of voice processing deficits in AD have not been systematically investigated to date, although several of these studies speculated impairments were the result of posterior cortical dysfunction.

In typical AD perisylvian disease is found early in the disease (Minoshima, Giordani, Berent *et al.* 1997; Neary, Snowden, Shields *et al.* 1987) affecting posterior temporal regions implicated in the perceptual processing of auditory objects. Despite anatomical overlap with PNFA, deficits of language production, grammatical and phonological processing are uncommon in typical AD. Instead abnormalities in temporoparietal cortex have been associated with other cognitive impairments, in particular with a cluster of symptoms that include deficits on language, spelling, calculation, and working memory tasks (Stopford *et al.*2008). As deficits in AD can affect more posterior cortical regions in the parietal lobe, AD may be a useful disease to investigate the roles of temporal and parietal regions implicated in phonagnosia (Van Lancker *et al.*1989).

Summary

Typical AD primarily presents with memory difficulties, with other cognitive domains including visuo-spatial skills, language and executive functions, affected at presentation or with disease progression. Associations between voice processing and other patterns of deficit in AD are of interest to identification of different clinical phenotypes. Impairments on semantic memory tasks have been described in AD including impairments of familiar face and name recognition, however it is unclear whether these result from generalized degradation of semantic knowledge stores (as predicted in SD), semantic access impairments, or whether they are underpinned by perceptual deficits. Familiar voice recognition has not been tested in AD to date. Recognition of

prosody and non vocal auditory objects, including environmental sounds and musical melodies, have been found to be affected early in the disease; however the neuroanatomical correlates of such deficits have not been reported. It is likely that temporo-parietal atrophy may affect regions critical to the perceptual analysis of voices, however very few studies have investigated perceptual levels of auditory or voice processing. Investigating performance on perceptual and associative voice processing tasks in AD and their neural correlates may further understanding of an under-recognised class of symptoms in this disease.

1.9. Aims of this Thesis

The studies in this thesis aim to delineate voice and accent processing mechanisms in neurodegenerative disease using neuropsychological tests and to explore the relations between voice processing test scores and grey matter volume in patients with dementia syndromes using VBM. The voice processing studies aim to evaluate current cognitive models which implicate separable perceptual and semantic voice processing stages and to investigate the neuroanatomical substrates for these processing stages. A further aim is to investigate vocal semantic processing mechanisms in relation to both semantic analysis of person knowledge in other modalities, and semantic analysis of other auditory objects, such as environmental sounds and music. The neurocognitive bases of two aspects of processing of accents are investigated in this thesis: accent recognition and comprehension of accented speech using novel neuropsychological assessments and VBM. Regional and foreign accent processing has rarely been investigated in lesion cases or in neurodegenerative disease, and is explored here in the context of voice processing models and auditory processing deficits in the target diseases.

The key focus of this work is to assess voice processing in neurodegenerative disorders with anatomically relevant frontal, temporal and parietal pathology, including typical AD and three syndromes of FTLD: bvFTD, SD, and PNFA. A growing body of functional imaging studies suggests temporal, frontal and parietal cortices are involved in perception of voices, recognition of familiar voices, and foreign accent discrimination. A joint behavioural and neuroanatomical approach is used to investigate the cortical organisation of voice and accent processing and to understand the brain basis of important nonverbal symptoms in these neurodegenerative diseases.

Specific aims

The specific aims of this thesis are summarised as follows:

1. To delineate voice processing deficits in the target diseases bvFTD, SD, PNFA and AD using neuropsychological tests designed to assess current cognitive models of voice processing that implicate separable perceptual and semantic voice processing mechanisms.

2. To demonstrate the brain basis of these deficits using VBM.

3. To further investigate semantic voice processing mechanisms in relation to semantic analysis of other complex auditory objects and to semantic processing of person knowledge in other modalities, by comparison to face and name processing.

4. To use novel neuropsychological tests to detect deficits of accent recognition and comprehension of accented speech in the target neurodegenerative syndromes of PNFA and AD.

5. To explore the neural bases of any deficits of accent processing using VBM.

6. To relate the above findings to patterns of nonverbal symptomatology in the target diseases.

1.10. Chapter Outline & hypotheses

Study 1 (Chapter 3)

Models of both voice processing and more broadly auditory object processing agree broadly on hierarchical processing of person information from early perceptual to higher semantic levels of processing. Key unresolved issues include the degree of modality-specificity of voice and face processing deficits; the level at which any modality specificity arises; the extent to which perceptual and semantic levels of processing are interdependent; and the status of voices versus other categories of auditory objects, and other fine-grained semantic categories beyond the domain of person knowledge. These issues are addressed in a case control study of two patients with deficits of person knowledge. The index case, patient QR, with bvFTD exhibited progressive loss of recognition of familiar voices as a leading clinical symptom, while the second patient, KL, presented with progressive prosopagnosia without a clinical complaint of altered voice recognition.

Selective deficits for voice recognition (relative to face and name modalities) could present either as associative or apperceptive impairments. According to models of nonverbal auditory object processing, low level perceptual deficits would result in a generalised auditory agnosia.

Apperceptive deficits are likely to result in impaired processing and recognition of other auditory objects, in particular those requiring high level timbral feature analysis such as musical instruments, but may not extend to deficits in recognition of all other auditory objects which are less reliant on timbre processing such as particular environmental sounds. These aspects are specifically addressed in this experiment.

Study 2 (Chapter 4)

In this study the neuropsychological and neuroanatomical signatures of voice processing are investigated in a group study of two canonical dementias, tvFTLD and AD, syndromes affecting temporal and parietal regions implicated in voice processing. Processing of voices is assessed in relation to current models of voice processing implicating perceptual and semantic mechanisms, and compared to processing of faces and names, in order to assess the modality- and material-specificity of any voice processing deficit.

Group level neuroanatomical correlates of voice processing performance are assessed using VBM and considered in relation to the neural correlates of voice processing implicated in studies of phonagnosia in brain-damaged cases and functional imaging studies of healthy controls. Distinct profiles of phonagnosia in tvFTLD and AD are hypothesized, with more severe associative impairment in tvFTLD and relatively more prominent apperceptive impairment in AD. It is further hypothesised, based on anatomical evidence in the healthy brain, that semantic deficits in processing voices would be associated with atrophy of ATL regions, and voice apperceptive deficits are associated with atrophy of more posterior temporo-parietal regions.

Study 3 (Chapter 5)

In this study the cognitive and neuroanatomical bases of accent processing are investigated in a group study of PNFA and AD. A novel neuropsychological battery is used to assess these cognitively impaired patient groups, addressing two aspects of accent processing: the intelligibility of accented speech (accent comprehension) and recognition of non-native regional and foreign accents (accent recognition). Neuroanatomical associations of behavioural performance are assessed using VBM. It is hypothesised that these dementia syndromes are associated with separable behavioural profiles of impaired accent processing. It is further hypothesized that accent comprehension and accent recognition performance have overlapping neuroanatomical associations in the postero-lateral and anterior temporal lobe regions previously shown to be critical for other aspects of vocal signal processing.

2. General Methods

2.1. Subject characterisation

All studies reported in this thesis were approved by the local institutional research ethics committee and all subjects gave informed consent in accord with the principles of the Declaration of Helsinki. Further characterisation of the subjects assessed in each experiment will be described in Study 1 (Chapter 3), Study 2 (Chapter 4) and Study 3 (Chapter 5).

2.1.1. Patients

Deficits of voice recognition were addressed in the context of frontotemporal lobar degeneration (FTLD) and Alzheimer's disease (AD). All patients were recruited from the tertiary Cognitive Disorders Clinic at the National Hospital for Neurology and Neurosurgery. Experimental tasks were administered to patients with a diagnosis of one of three variants of FTLD according to the consensus criteria of (Neary *et al.*1998): bvFTD, SD and PNFA. Patients would have also fulfilled recent criteria for probable bvFTD (Rascovsky, Hodges, Knopman *et al.* 2011) and Primary Progressive Aphasia (PPA) (Gorno-Tempini, Hillis, Weintraub *et al.* 2011), on the basis that all patients had supportive MRI. Different FTLD subjects were tested in each of the three studies, with the exception of one subject (KH) recruited as a case control in Study 1 (Chapter 3) who was also used as part of the tvFTLD patient group in Study 2 (Chapter 4).

Patients with a clinical syndrome of typical mild to moderate AD led by memory decline were recruited. Twenty two subjects fulfilling modified NINCDS-ADRDA (National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association) criteria for a clinical diagnosis of probable AD (Dubois, Feldman, Jacova *et al.* 2007) were recruited into Study 2 (Chapter 4), twenty of these patients also completed tests for Study 3 (Chapter 5).

2.1.2. Controls

Experimental tasks were also administered to a control group of healthy older individuals recruited from a database of control subjects through previous participation in research studies at the Dementia Research Centre and local community organisations. Controls were all native British residents with no history of neurological or psychiatric illness.

Twenty-four controls were recruited for Study 1, an additional 11 controls were recruited for Studies 2 and 3 (N=35). A separate group of controls were used for the pilot study described in Appendix A.1.

2.2. Background measures

2.2.1. General neuropsychological assessment

Patients and healthy control subjects completed general neuropsychological assessments including tests of IQ, executive function, auditory verbal working memory, object perception, single word comprehension and naming. Subjects were also assessed on standard tests assessing identification of faces (Study 1) and topographical landmarks (Studies 1 and 2); examples of 'unique entities' in the visual modality. A full list of neuropsychological tests utilised in this thesis is displayed in Appendix A.2. Individual subjects' neuropsychological test scores were defined as impaired if performance fell below the 5th percentile of the standardisation sample.

2.2.2. Assessment of peripheral hearing

Presbycusis or peripheral hearing loss is common in older adults and typically involves loss of higher frequency hearing. A screening peripheral hearing test was used in all subjects to assess the magnitude of any hearing loss and potential influences on auditory voice cognitive tests. A pure tone audiometry test involved a procedure adapted from a commercial screening audiometry software package (AUDIO-CD^{TM}, http://www.digital-recordings.com/audiocd/audio.html). A notebook computer was used to administer tones (via headphones) and to record responses; the test was conducted in a quiet room. Five frequency levels (0.5, 1, 2, 3, 4 kilohertz (KHz)) were assessed: at each frequency, subjects were presented with a continuous tone that slowly and linearly increased in intensity (1 decibel (dB) per second (sec)). Subjects were instructed to tap as soon as they could detect the tone and the response time was stored for offline analysis. The mean value of response time (i.e., detection threshold) for three presentations of the same tone in the right ear was taken as the detection threshold for that frequency.

In Studies 2 and 3 differences in mean response time between the groups for each pure tone frequency adjusted for age and gender are reported, with p values found with z-tests using bootstrap (2000 bootstrap samples) standard errors. In Study 1, subjects' response times were compared to age determined normal limits provided as part of the software. In the two cases

described in Study 1 in addition to screening hearing test, full audiometry was assessed clinically by audiologists to detect peripheral hearing loss. Any patient or control subject in Studies 2 and 3 reporting abnormalities of hearing also received a full audiometry assessment.

2.2.3. Assessment of media exposure

The background media exposure of the control subjects and patients was assessed formally in Studies 1 and 2 as a potentially relevant factor influencing familiar voice recognition ability. The results of the pilot study (described in further detail in Appendix A.1.) found that older adult control voice recognition performance on all measures (familiarity, naming and identification) were significantly associated with questionnaire measures of media exposure news exposure and TV (television) watching (see Appendix A.1.) but not other demographic measures of the controls. A brief questionnaire recording the estimated average number of hours spent each week watching television and listening to the radio and the average number of news exposures each week (over the period of previous three months) was completed by control subjects and by patients' carers.

2.3. Experimental investigations of voice and accent processing: plan and general procedure

Auditory stimuli were presented from digital wavefiles on a laptop computer at a comfortable listening level (typically 70-80 dB) in a quiet room, in Studies 2 and 3 stimuli were delivered through headphones, in Study 1, voice stimuli were delivered in free-field at a comfortable constant listening level. Visual stimuli were presented as clear high-quality black and white photographs. Verbal stimuli were simultaneously presented as written words and spoken by the examiner (control subjects were presented with written words only). The tests were presented in a fixed order to all participants; within a test, the order of stimuli was randomised, but the same for all subjects. 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 was given about performance during the test. Voice and accent 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. The experimental tests were administered to subjects over several sessions.

2.4. Experimental investigations: Perceptual analysis of voice attributes

Unlike receptive tests of linguistic processing there is no gold standard for assessing paralinguistic voice perceptual processes. In this thesis, low level perceptual tasks were selected

with several constraints: that they used vocal stimuli to ensure that assessments were not simply testing aspects of auditory perception that were irrelevant to voice processing, secondly that they should be ecologically valid tasks, lastly that there should be evidence for their validity from psychological and neuroimaging studies of normal healthy volunteers. Perceptual encoding of two vocal attributes which serve as important cues to identity were selected: judgments of vocal size and gender. These are tasks that require intact processing of vocal pitch and aspects of timbre, but enable a listener to extract useful invariant information about the speaker. The brain mechanisms that represent these attributes may be at least partly voice-specific (Belin *et al.*2000; von Kriegstein *et al.*2006). To date there have been no previous neuropsychological studies of impaired gender or vocal size discrimination in brain-damaged subjects, in accord with the lack of descriptions of apperceptive phonagnosia in the literature.

The ability to recognise gender from the voice is a natural and easy task for the healthy brain; normal listeners are able to discriminate gender to levels of over 90% accuracy even in degraded listening conditions such as reduced spectral resolution (Childers *et al.*1991; Remez *et al.*1997; Wu & Childers 1991). Infants are able to categorize the gender of voices from 8 months of age (Patterson & Werker 2002), suggesting that the perceptual cues underlying gender can be easily perceived. It is generally believed that the perception of pitch determined by F_0 (or Glottal Pulse Rate) is the strongest cue to gender, as there is little overlap in the F_0 between males and females: female voices are typically an octave higher than males, as described in Section 1.2.2. Formant frequencies are also lower in males compared to females as a result of differences in vocal tract length (VTL), and are likely to influence gender discrimination (Childers *et al.*1991; Smith *et al.*2005; Smith, Patterson, Turner *et al.* 2005; Smith *et al.*2007). The neural correlates of gender discrimination performance are likely to overlap with the neural correlates of pitch perception in primary and non- primary auditory association cortices (posterior superior temporal cortices) (Kawahara & Irino 2004; Patterson *et al.*2002). Characteristics of vocal quality (timbre and prosody) are also proposed to differ between genders, for example a more “breathy” and melodic quality in females (Childers *et al.*1991; Klatt *et al.*1990; Singh *et al.*1978), however these are much more subtle cues for gender and likely to be subsidiary to pitch cues under normal listening conditions.

Vocal size is a fundamental and invariant perceptual property of voices which can be reliably extracted from the voice, and is likely to be encoded earlier than representations needed in voice discrimination (Smith *et al.*2005; Smith *et al.*2007; Vestergaard, Haden, Shtyrov *et al.* 2009; von

Kriegstein, Smith, Patterson *et al.* 2007). Perceptual judgment of speaker size is a fundamental task of auditory cognition in humans and other species, and VTL is an important cue for perception of speaker size by normal subjects; differences in VTL show systematic differences between children and adults, and between males and females. VTL not only determines the pitch of the formant frequencies (as described in Section 1.2.2, longer VTLs lower the frequency of formants) but different VTLs also lead to systematic differences in vocal timbre, as longer VTLs result in a greater presence of lower frequencies in the sound and longer decay times (for example resulting in a deeper lower voice in adult males). A computational algorithm has been developed to mimic these changes in timbre at any given pitch (Kawahara *et al.*2004), and studies using generated stimuli have shown that normal healthy volunteers are able to discriminate this characteristic of speaker size with ease (Smith *et al.*2005; Smith *et al.*2005). Healthy listeners are also able to discriminate auditory size for other auditory objects such as musical instruments, however, imaging studies implicate specialized mechanisms for the representation of the auditory size in speech formants in left posterior STG (von Kriegstein *et al.*2007; von Kriegstein *et al.*2006).

2.4.1. Tests of vocal gender and size

Vocal size and gender were firstly assessed in two subtests based on forced-choice responses. Subjects' ability to assign gender to vocal samples was based on all available auditory cues. Vocal samples (each 5 sec in duration) were derived from publicly available sources. 24 trials (12 male, 12 female) were presented; the task on each trial was to decide if the voice was male or female.

Categorical ('big' versus 'small') perceptual judgements of vocal size were assessed by manipulating VTL information in isolation, as previously described in normal subjects (Ives *et al.*2005). Stimuli were based on consonant-vowel syllables recorded by a single male speaker and digitally resynthesised using a previously described algorithm (Kawahara *et al.*2004) that allows apparent VTL to be varied independently of glottal pulse rate (voice pitch). Each syllable was presented at two extreme VTL values, one corresponding to a speaker height of 2 meters (equivalent to a very tall male, 'big') and the other to a height of 0.5 meters (equivalent to a child, 'small'), and randomly assigned one of four pitches within the normal human male vocal range (Study 1: 116, 120, 138, 158 Hertz (Hz); Study 2: 116, 120, 172, 190 Hz), which was varied independently of VTL. Equal numbers of "big" and "small" trials were randomly presented, on each trial subjects heard a sequence consisting of repetitions of the same stimulus: three

repetitions were presented in Study 1; four repetitions were presented in Study 2. Subjects were asked to judge if the sounds were made by a big person or a small person. 32 trials were presented in Study 1, whereas in Study 2, 20 trials were administered.

2.5. Experimental investigations: Voice discrimination

Beyond early perceptual encoding but prior to the attribution of meaning it is likely that voice processing entails an interposed stage of representation of the voice as a complex auditory object (Griffiths *et al.*2004; Warren *et al.*2006). This apperceptive stage of vocal processing can be assessed by tasks requiring discrimination of unfamiliar speakers, an auditory analogue of apperceptive processing in the visual domain (described in Section 1.4.2). Several studies have assessed "apperceptive" voice processing, and have mostly assessed the ability to discriminate between unfamiliar voices of the same gender: requiring listeners to decide if two audio samples played in succession are the "same" or "different" speakers. Studies of phonagnosia have shown a double dissociation between performance on this test and recognition of familiar voices (Garrido *et al.*2009; Van Lancker *et al.*1987; Van Lancker *et al.*1985). This task, in theory cannot be performed solely using low level cues between voices (such as pitch, gender or size information) and is likely to represent complex or later stage of perceptual processing. It is likely that such fine-grained analysis of voices is dependent on both pitch and timbre processing (Belin *et al.*2004).

2.5.1. Tests of speaker discrimination

Previous studies of speaker discrimination have used voice samples from foreign speakers (Garrido *et al.*2009) or samples that are fairly long in duration, such as utterances of a sentence spoken by two speakers in succession (Van Lancker *et al.*1987; Van Lancker *et al.*1988; Van Lancker *et al.*1989). Here, a novel voice discrimination task was created in which subjects were required to detect a change in speaker within a short spoken phrase using native speakers. The verbal content of the phrase was highly over-learned spoken phrases: comprising days of the week 'Monday, Tuesday, Wednesday, Thursday' or months of the year 'January, February, March, April'. In order to control for gender, age and accent factors, all speakers were female, aged 21–31 years, with a standard Southern English accent. Recorded single words were concatenated with fixed inter-word gaps (0.1 sec) to equate overall speech rate. If the sequence contained a speaker change, this change always occurred at the midpoint of the phrase, to maximise available vocal information for each speaker. If the sequence was spoken by the same

speaker, recordings from two separate recording sessions were used, and stimuli were presented also with a change in recording at the midpoint of the phrase.

Two versions of the test were created. In the original 'difficult' version of the test (used in Studies 1 and 2), inter-speaker variations in vocal pitch were controlled by setting f_0 of recorded stimuli at 220 Hz using Goldwave® software. An 'easy' version of the task was also created in which voice pitch (f0) was not fixed. In Study 1, 48 'difficult' items were administered: 24 trials (12 speaker fixed, 12 speaker change) consisting of spoken sequences of days of the week were presented, followed by 24 trials (12 speaker fixed, 12 speaker change) using sequences of months. In Study 2 all stimuli presented used spoken sequences of days of the week: first 28 trials of the 'easy' discrimination items were administered (14 speaker fixed, 14 speaker change), followed by 12 items from the 'difficult' speaker discrimination test (6 speaker fixed, 6 speaker change). On each trial, 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, Hamsher, Varney *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 were normalised for age and education and scored out of 56.

2.6. Experimental investigations: Voice recognition

In order to investigate subjects' ability to recognise familiar voices, audio recordings of famous personalities were used. There may be some differences between neurocognitive representations of famous people compared to personally familiar people (Giovannetti, Sestito, Libon *et al.* 2006; Joubert, Mauries, Barbeau *et al.* 2004; Snowden, Griffiths, & Neary 1994), however recognition of famous individuals using photos of faces is an established neuropsychological paradigm for detecting prosopagnosia (Warrington *et al.*1967), and has been used in the voice modality to detect phonagnosia in brain-damaged subjects (Ellis *et al.*1989; Gainotti *et al.*2003; Hanley *et al.*1989; Joubert *et al.*2006; Neuner *et al.*2000; Van Lancker *et al.*1982; Van Lancker *et al.*1988; Van Lancker *et al.*1985). In order to assess the modality specificity of any deficit, aspects of semantic processing of famous voices (voice recognition) were compared to recognition of faces and names.

Previous work has shown that normal individuals have greater difficulty recognising public figures from voice than from faces or name (Damjanovic *et al.*2007; Ellis *et al.*1997; Hanley *et al.*1998). Not only are voices not as easily recognised as faces, a bias also exists in recognition of public figures' faces versus their names: famous faces are less well recognised than names in normal controls and amnesic AD (Young, McWeeny, Hay *et al.* 1986). The intrinsic relative efficiency of recognition from voice versus face information is an important issue when comparing voice recognition to person identification from different modalities. Poor normal control performance on voice recognition tests is potentially a serious limitation on the characterisation of any deficit; in a recent study of developmental phonagnosia controls were excluded due to having insufficient television exposure (Garrido *et al.*2009).

Pre-morbid familiarity with a set of voices has been assured by using the voices of people personally familiar to subjects in neuropsychological and neuroimaging studies (Gentileschi *et al.*1999; Gentileschi *et al.*2001; Imaizumi *et al.*1997; Nakamura *et al.*2001; von Kriegstein *et al.*2006; von Kriegstein *et al.*2005), however the use of such stimuli complicates attempts to quantify performance between patients and in relation to controls. Addressing the ability of older adults to recognise public figures is necessary to effectively quantify dementia patient performance; older adults are likely to differ from younger adults in their exposure to media personalities and in their ability to perform auditory tasks due to high frequency hearing loss due to ageing (see Section 2.2.2) or differences in auditory expertise (Hailstone *et al.*2009; Halpern, Bartlett, & Dowling 1995). Normal voice and face recognition ability was quantified in a pilot study in healthy older adult controls, further details of methods and results are provided in Appendix A.1.

Supporting previous work (Damjanovic *et al.*2007; Ellis *et al.*1997; Hanley *et al.*1998; Neuner *et al.*2000) the pilot study found that across familiarity, naming and identification semantic tasks, famous individuals were consistently better recognised by healthy controls in the face modality compared to from their voice. This may be partially due to the higher frequency of exposure to the faces of public figures in the media (for example in the news) compared to voices, alternatively as proposed by Hanley and colleagues, this may represent systematic differences in access to person-specific information between the two modalities (Hanley & Damjanovic 2009; Hanley *et al.*2000). The results of the pilot study do not resolve this issue, but show that matching the control recognition frequency across modalities is likely to be difficult (as previously indicated in studies

in which face stimuli were presented blurred in order to match control recognition frequency to voices (Hanley *et al.*2009; Hanley *et al.*2000)); nearly all public figures assessed in the pilot study were recognised by more controls from their face than voice (reported in Appendix A.1). Familiarity and frequency of exposure are likely to be important constraints on the robustness of semantic representations to neurodegenerative disease (Lambon Ralph *et al.*1999; Lambon Ralph *et al.*2001; Lambon Ralph, Patterson, Garrard *et al.* 2003), however according to person recognition models the same individual represented in different modalities should activate the same multimodal perceptual representation (PIN) and access the same amodal semantic representations associated with the person. In this thesis, instead of matching control recognition frequency across modalities, identical famous individuals were used in all modalities both to control semantic factors (activation of PIN and amodal representations) but also, here interest was not just differences between modalities but the profile of overall patient performance relative to healthy controls across modalities.

As for other kinds of person knowledge, semantic processing of voices leading to identification could in principle comprise a number of different subprocesses, and the relations between these have not been defined. For the purposes of this study separate semantic subtests were designed based on familiarity judgement, free identification (by naming or verbal biographical description), and forced-choice cross-modal recognition. The specificity of any voice recognition deficit was assessed by testing recognition of the same set of famous people represented in two other modalities: faces and names. For famous faces, the same four semantic tasks were used as for voices, for famous names, familiarity judgement and biographical description were used. As voice recognition was the primary focus here, voice tasks (familiarity, naming, biographical description) were presented first, followed by face tasks (familiarity, naming, biographical description), and name tasks (familiarity, biographical description). Finally, cross-modal recognition of voices and faces was assessed; in order to avoid priming effects on face recognition during the voice recognition task, face recognition was assessed first (matching a face to a choice of names), followed by voice recognition (matching a voice to a choice of faces-name pairs).

2.6.1. Tests of familiarity of voices, faces and names

In this subtest familiarity judgments on famous voices were compared with familiarity judgments on faces and names for the same famous individuals. 24 public figures whose voices were best identified from voice by pilot study controls were selected (correctly identified by 64-92% of

controls (mean = 75.0, SD=9.0)). These individuals (see Appendix A.3) comprised ten politicians, five actors, seven media personalities from television and radio, and two members of the British royal family. Face photographs and written and spoken names of the same set of 24 famous individuals were used for face and name processing tasks, respectively.

The set of 24 famous voices was supplemented by 24 unfamiliar voices and faces (as classified by over 75% of healthy controls in the pilot study (Appendix A.1)) which were 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 on each trial was to decide if the stimulus was familiar in a forced choice ('yes – no') protocol.

2.6.2. Tests of voice, face and name identification

This subtest assessed subjects' ability to name or to identify the set of 24 public figures (described above in Section 2.6.1) by providing other biographical details (e.g., an event closely associated with the person, occupational information), in line with the criteria used by Snowden and colleagues (Snowden *et al.*2004). In voice and face modalities, on each trial, the task was to identify the person as precisely as possible; if the subject was not able to name the person they were encouraged to provide other information about them. In the name modality, on each trial the subject was required to provide identifying information about the person. For voice stimuli, national or regional origin was not accepted as evidence of person recognition, since this could be based on accent cues alone.

2.6.3. Tests of cross-modal recognition of voices and faces

SD patients often perform poorly on verbal retrieval tasks; accordingly, a cross-modal matching task was employed 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 public figures; 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 organisational principle in the domain of person knowledge: (Crutch & Warrington 2004)). The set of 24 faces was presented first, and the task on each trial 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 on each trial was to match the voice to one of the face – name pairs.

2.7. Experimental investigations: Accent processing

Neuropsychological assessments of accent processing are described in Chapter 5, Sections 5.2 and 5.3.

2.8. Neuroimaging: Structural MRI in dementia

High resolution brain MRI images are frequently used to measure pathological changes in dementia because of the greater anatomical detail compared to images from other methods; three-dimensional volumes and strong grey and white matter contrast resolution result from visualisation of the radio signal produced by different tissue types. For example prominent atrophy of the temporal lobes in semantic dementia is well visualized by high-resolution MRI, but may not be detected by CT (computerised tomography). As a non-invasive method for visualizing and quantifying atrophy in different structures and for measuring rates of progression of atrophy, MRI is frequently used in neurodegenerative diseases for purposes of diagnosis and to monitor rates of disease progression.

It is possible to quantify the volume of different structures either by manually outlining healthy and diseased tissue or regions of interest (ROIs), or a template is used to measure healthy and diseased tissue within this template region. Healthy or atrophic regional volumes are then used in statistical analyses to calculate either population volumes or to correlate regional GMV with scores on neuropsychological tests. Alternative approaches have also been used to quantify and locate pathological differences between patient groups including VBM or cortical thickness methods which examine tissue types across the brain (in particular, grey or white matter) to identify regions where statistically significant differences in image intensities exist between groups. The advantages of these methods are that they are semi-automated and unbiased, because they do not require a priori hypotheses about particular ROIs. In this thesis patterns of atrophy on MRI on the basis of radiological visual descriptions were utilised in descriptions of individual cases in Study 1 and group patterns of atrophy were analysed with VBM in Studies 2 and 3 to correlate patterns of atrophy with neuropsychological scores as described below.

2.8.1. Structural image acquisition

For all patients willing to volunteer and tolerate a scan, T1-weighted volumetric MR images were acquired on a Siemens Trio TIM 3 tesla (3T) scanner (Siemens Medical Systems, Erlangen, Germany). 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 x 256 acquisition matrix, giving approximately isotropic 1.1 mm cubic voxels.

2.9. Voxel-based morphometry (VBM)

VBM involves aligning patients' volumetric MRI scans into the same spatial framework so that groups of scans can be statistically compared on a voxel-by-voxel (point by point) basis. Statistical analysis is frequently performed to localize group differences in patterns of atrophy in dementia (Henley, Wild, Hobbs *et al.* 2009; Lehmann *et al.*2009), and more recently methods have been developed to relate atrophy with neuropsychological test scores. Traditional methods for studying lesion-deficit relationships required that groups of patients were prespecified either on the basis of the presence or absence of a behavioural deficit (Adolphs, Damasio, Tranel *et al.* 2000) or on the basis of their lesion location (Van Lancker *et al.*1982; Van Lancker *et al.*1989), which meant that a large number of patients were needed and variation between patterns of atrophy or in test scores were discarded. In VBM smaller numbers of patients can be used, patterns between and across patient groups can be analysed, and variations in scores are utilised in the analyses.

2.9.1. VBM image processing

VBM relies on successful preprocessing of the scans to ensure anatomical correspondence between all the brain images, and modifications to methods are desirable when studying atrophic or lesioned brains. Here, MR brain images were processed using MATLAB 7.2 (The MathWorks, Inc., Natick, MA, USA) and an image analysis package SPM8 software (Statistical Parametric Mapping, Version 8; http://www.fil.ion.ucl.ac.uk/spm) with default settings for all parameters. In preprocessing, raw brain images were first manually rigidly reoriented to standard space (the international consortium for brain mapping template). A series of steps were then performed using SPM; firstly the reoriented scans were segmented into grey and white matter using the new Segmentation Toolbox in SPM8. "Imported" grey and white matter segmentations were used in DARTEL (Differomorphic Anatomical Registration Through Exponentiated Lie Algebra) (Ashburner 2007), a toolbox for SPM8 which uses an algorithm for improved image registration between individuals which iteratively registers the segments to an evolving estimate of their

group-wise average (Ashburner & Friston 2009) resulting in improved localisation and relative to previous spatial normalization methods (Ashburner & Friston 2000). Grey matter segments were then normalised using the final DARTEL transformations and modulated to account for local volume changes (Ashburner *et al.*2000). Finally, the images were smoothed so that each voxel is made an average of itself and the surrounding voxels (the larger the size of the kernel the larger the surrounding region included), using an isotropic Gaussian smoothing kernel (Study 2: 8 mm full-width-at-half-maximum; Study 3: 6mm full-width-at-half-maximum). Smoothing helps to compensate for the inexact nature of the spatial normalization process, and also makes the data more normally distributed which increases the validity of parametric statistical analyses, and is based on Gaussian random field theory (Ashburner *et al.*2000).

2.9.2. VBM analyses

For each experimental test, associations between regional grey matter volume (GMV) and experimental test performance were assessed across and within disease groups using linear regression models. Total brain volume in healthy controls has been found to vary with head size (Acer, Sahin, Bas *et al.* 2007) and gender (Good, Johnsrude, Ashburner *et al.* 2001) and it has been shown that adjusting whole brain volume for total intracranial volume (TIV) can eliminate differences due to head size and gender (Whitwell, Crum, Watt *et al.* 2001). TIV is the volume within the cranium which includes the brain, meninges and cerebrospinal fluid (CSF). It was measured outside SPM on each subject's T1 weighted segmented images using an algorithm to summate and linearly interpolate the sum of the volume of grey matter, white matter and CSF segmentations (Whitwell *et al.*2001). For each experimental subtest, GMV was modelled as a function of score, including age and TIV and group as covariates. A score-by-group interaction term was included in the combined-groups analyses to allow different GMV-score slopes between groups. Further analyses were carried out in Studies 2 and 3 (described in Section 4.2.5 and Section 5.5, respectively).

An explicit analysis mask was used to exclude any voxels for which more than 20% of the images had an intensity value of less than 0.1. This proportional thresholding procedure has been shown to improve visualisation of markedly atrophic brain regions compared with the default "absolute thresholding" mask option in SPM(Ridgway, Omar, Ourselin *et al.* 2009). For each experimental test, grey matter associations were assessed both over the whole-brain and within the ROI specified by our prior anatomical hypotheses. A voxel-wise statistical threshold $p < 0.05$ family-wise-error (FWE)-corrected for multiple comparisons was applied in all analyses (a global $p<0.05$

FWE-corrected threshold was applied in the combined-modalities conjunction analysis). Small volumes for were created manually for each study separately in MRIcron® (http://www.cabiatl.com/mricro/mricron/index.html) from a study-specific template created by warping all native space whole-brain images to the final DARTEL template and calculating the average of the warped brain images. Separate small volumes were created for the right and left hemispheres and each was intentionally generous to ensure adequate coverage of the whole temporal lobe, anterior to the temporo-occipital junction, the temporo-parietal junction and inferior parietal lobe. These regions were selected because temporal and inferior parietal regions have been previously been implicated in vocal processing (Chapter 1) (Adank *et al.*2009; Belin *et al.*2003; Belin *et al.*2002; Hanley *et al.*1989; Joubert *et al.*2006; Lewis *et al.*2009; Rohrer *et al.*2010; Scott *et al.*2006; Van Lancker *et al.*1988; von Kriegstein *et al.*2004; Warren *et al.*2006). All attributions within each small volume were subsequently inspected to ensure anatomical accuracy. Statistical parametric maps (SPMs) were displayed as overlays on the study-specific template.

In order to report coordinates of local maxima in the standard stereotactic MNI (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 SPM, and the DARTEL coordinates were transformed using the estimated affine mapping to MNI space.

2.10. Statistical analyses of behavioural data

Statistical analyses in Studies 2 and 3 were carried out in STATA release 9.2 (Stata Corporation, College Station, Texas, USA). Fisher's exact test was used to assess group differences in gender, for all other demographic variables differences between the groups on demographic measures, neuropsychological and experimental test scores were assessed using z-tests and 95% Wald type confidence intervals, with standard errors calculated using bootstrapping (2000 replicates). Further details of statistical analyses of group behavioural data in Studies 2 and 3 are described in Chapters 4 (Section 4.2.4) and 5 (Section 5.4).

In Study 1, single-case results were assessed in relation to the control sample using a method designed for use with a small control sample in which the control sample statistics are treated as statistics rather than as population parameters, and are compared using a t-test to the single case statistic (which is treated as a sample of N=1 (Crawford & Howell 1998)). The advantage of this

method is that it has been shown to control the Type 1 error rate regardless of the control sample size, whereas the z test has been shown to be vulnerable to a Type 1 error with smaller control sample sizes (Crawford & Garthwaite 2005; Crawford *et al.*1998). Methods have been developed for comparing the difference between a patient's performance on two or more tasks which control the Type 1 error rate by the same authors, including the Revised Standardized Difference test. In Study 2, this test was used to compare individual patient performance on scores between two modalities of presentation relative to differences in score in the control sample (Crawford *et al.*2005). In the single cases in Study 1, McNemar's test was used to compare performance between modalities on corresponding items between items (McNemar 1947).

For both group Studies 2 and 3, individual subject performance was classed as impaired if below the 5^{th} percentile cut-off score for the healthy control group.

3. Study 1: Progressive associative phonagnosia: a neuropsychological analysis

3.1. Introduction

In the first study in this thesis voice processing performance was addressed in two patients with deficits of person knowledge in the context of FTLD. The index case, patient QR, exhibited progressive loss of recognition of familiar voices as a leading clinical symptom of bvFTD, while, KL, a case with a diagnosis of tvFTLD with progressive right temporal lobe atrophy presented with progressive prosopagnosia without a clinical complaint of altered voice recognition. Several issues raised by models of voice recognition and auditory object cognition were addressed in this case control study which are described below.

Cognitive models of voice recognition provide a framework for analysing voice processing: models agree broadly on the segregation of perceptual processing (via parallel processing of faces, voices and name stimuli), however detailed predictions of these models and their neuropsychological instantiation have yet to be fully worked out. In the few studies in degenerative disease in which associative phonagnosia has been described (primarily in progressive prosopagnosic cases) perceptual voice processing was not tested, and the extent to which perceptual and semantic levels of processing are independent is of interest to both these models and theories of auditory object processing. The degree of modality-specificity of voice processing impairment and the level at which any modality specificity arises are unresolved issues for voice processing models; associative voice recognition deficits have been described in degenerative cases, generally only after the onset of face recognition deficits (Evans *et al.*1995; Gainotti *et al.*2003; Gentileschi *et al.*1999; Gentileschi *et al.*2001; Joubert *et al.*2003) and therefore they may not be identified as a clinical issue (described in Section 1.8.1).

The extent to which processing of voices is separable from other complex non-verbal sounds has also not been established. Voice processing models predict that specialized auditory mechanisms are needed for voices, in particular at the level of high level perceptual analysis in the VRUs (Belin *et al.*2004) however there is little evidence neuropsychologically for dissociation at apperceptive levels of voice processing with other types of sounds, in particular it is likely that deficits co-occur with perception of sounds that are dependent on analysis of timbre (Goll *et*

*al.*2010; Mazzucchi *et al.*1982). There is evidence from studies of patients with focal lesions that phonagnosia can dissociate from agnosias for other sounds at associative levels of processing: in particular phonagnosia has been shown in a few cases with preserved environmental sounds recognition (Garrido *et al.*2009; Neuner *et al.*2000; Peretz *et al.*1994). However, impairments of voice processing are also described in cases with temporal lobe lesions combined with deficits of recognition of other auditory objects (Assal *et al.*1981; Peretz *et al.*1994), it is plausible therefore that the two may co-occur in degenerative disease. Recognition of voices in this study was compared with recognition of another category of auditory objects by probing identification of environmental sounds.

Selective deficits of voice processing versus other kinds of complex sounds may either reflect the privileged ecological status of human voices or rather the greater demands of processing unique auditory exemplars; voice identification requires fine-grained perceptual and semantic processing as a single highly differentiated category of complex sounds. Similar arguments have previously been advanced to challenge claims that human faces constitute a privileged category of visual objects (Gainotti *et al.*2008). The status of voice processing in comparison to other fine-grained semantic categories beyond the domain of person knowledge is of interest to auditory object processing models and theories of semantic processing, which make hypotheses about categorical segregation and/or levels of processing, for example hypothesising segregation of generic versus fine-grained knowledge representations in the inferior and anterior temporal lobes (Mion, Patterson, Acosta-Cabronero *et al.* 2010). This issue was addressed by testing recognition of an alternative highly differentiated category of complex sounds: musical instruments.

A further issue for cognitive models of voice processing is the extent to which vocal identity information dissociates from recognising vocal emotion (see Belin's model of voice processing in Section 1.7.2). The only study to have shown this neuropsychologically is developmental phonagnosic case KH (Garrido *et al.*2009) who displayed associative deficits of voice recognition but normal performance at recognising vocal emotion. Patients with FTLD (in particular bvFTD) often show altered responses to emotions in various input modalities, including the voice modality (Keane *et al.*2002; Snowden *et al.*2008), whereas in SD similar alterations in social and emotional behaviour can co-occur with impairments of nonverbal object recognition including people. In SD behavioural and person recognition symptoms have been associated with atrophy affecting the right ATL (Brambati *et al.*2009; Chan *et al.*2009; Edwards-Lee *et al.*1997; Perry *et al.*2001). In this study it was of interest to compare recognition of familiar identities to

recognition of emotions in order to assess whether these are cognitively and neuroanatomically separable processes in patients with FTLD.

3.2. Methods

3.2.1. Subject details

Patient QR

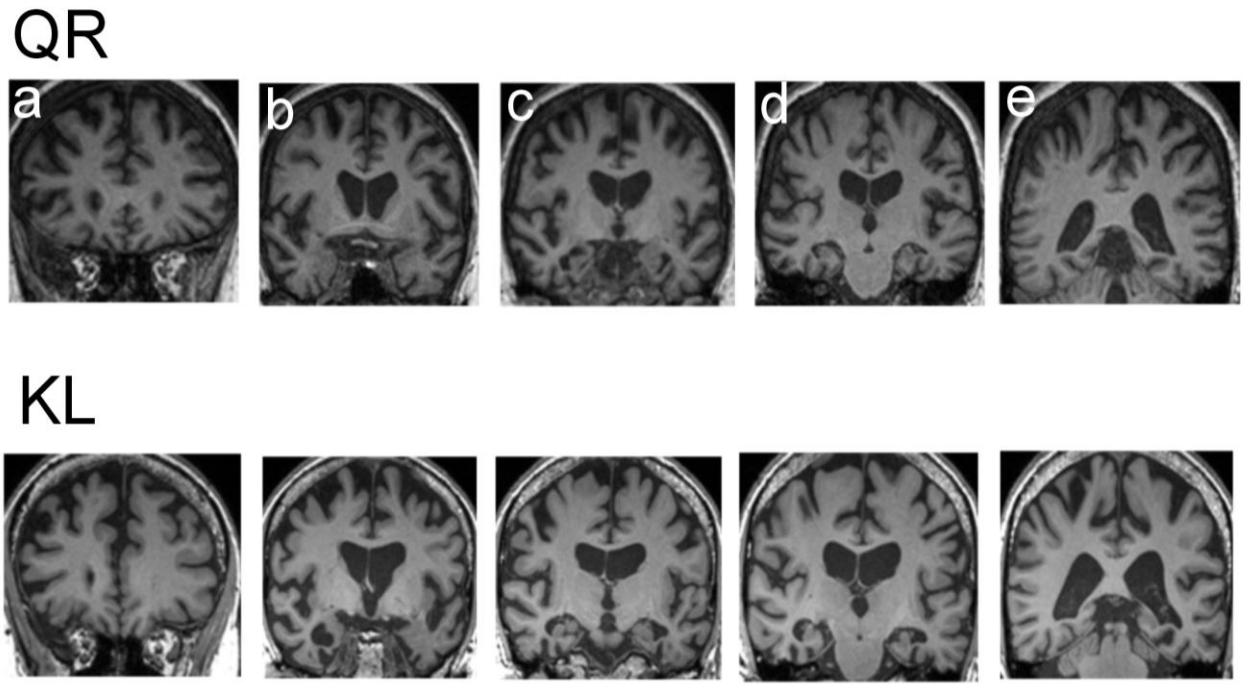
This 61-year-old right-handed female hairdresser presented with a two year history of insidiously progressive behavioural decline, with impassivity, obsessionality, clock-watching, loss of empathy and development of a sweet tooth. Impaired voice recognition was an early symptom. When first assessed she was no longer able to identify the voices of her children on the telephone, nor did she evince any sense that their voices were familiar. In contrast, recognition of faces had not been similarly affected: she consistently recognised family members, and despite the suggestion of some recent difficulty identifying friends in social situations, she continued to exhibit a sense of familiarity toward them. On examination there was evidence of executive dysfunction, disinhibition, perseveration and impulsivity. Naming and verbal memory were impaired whereas early visual perceptual skills were preserved. The general neurological examination was unremarkable. Peripheral hearing assessed using pure tone audiometry was within normal limits for age. Brain MRI (Figure 3.1) showed bilateral fronto-temporal atrophy somewhat accentuated in the right ATL but extending posteriorly within the temporal lobe and including the STS, with no significant cerebrovascular changes. The clinical diagnosis was bvFTD.

Patient KL

This 72-year-old left-handed male academic presented with an eight year history of insidious cognitive decline; initially reporting a difficulty recognising neighbours and other close acquaintances, followed by progressive difficulties with word finding and topographical memory and mild behavioural changes. He had been born in the US but had lived in the UK periodically for over 50 years and consistently for the last 11 years. There was no history to suggest phonagnosia though he reported that he found understanding unfamiliar accents increasingly difficult. On examination there was evidence of mild disinhibition and impaired recognition of famous faces, with preservation of early visual perceptual skills. The general neurological examination was unremarkable. Peripheral hearing assessed using pure tone audiometry was within normal limits for age. Brain MRI (Figure 3.1) showed bilateral predominantly ATL

atrophy, more marked on the right side and in the inferior temporal cortices including the fusiform gyrus. The clinical diagnosis was tvFTLD with progressive right temporal lobe atrophy.

Figure 3.1. Representative T1-weighted coronal brain MRI sections from each patient



The right hemisphere is shown on the left side of each image. Sections have been selected to show the following regions of potential relevance to voice processing deficits: a, frontal lobes; b, temporal poles; c, ATLs; d, mid-temporal lobes including Heschl's gyri; e, temporo-parietal junction. Focal cerebral atrophy is shown in both patients: in QR, bilateral fronto-temporal atrophy accentuated in the right ATL and extending posteriorly and including the STS; and in KL, bilateral predominantly ATL atrophy, more marked on the right side and in the inferior temporal cortices including the fusiform gyrus.

Healthy controls

Perceptual and semantic tasks were administered to 24 control subjects (17 female; mean age=64.5, SD=4.3, range: 55-73; mean years of education 15.5, SD=3.5, range 11-25). All had normal screening audiometry. Between 20 and 24 controls completed each of the voice processing tests, and 14 controls also completed a test of environmental sound recognition. In addition, a test of vocal emotion recognition was administered to a separate group of 22 older controls (12 female; mean age=67.2, SD=8.8, range: 53-78).

Background neuropsychological assessment

The performance of QR, KL and healthy controls on general neuropsychological tests and standard tests assessing identification of faces and topographical landmarks, examples of 'unique entities' in the visual modality, are summarised in Table 3.1.

Table 3.1. Summary of patient and control performance on background neuropsychological assessment

	QR score (percentile)	**KL score (percentile)**	**Controls mean (SD)**
General neuropsychological tests			**n=23**
NART Full Scale IQ	86	113	120.9 (6.3)
MMSE (/30)	28	25	n/d
WAIS III Digit span (forwards, back) (/14)	12,5 (60-80th)	14,5 (80-95th)	n/d
Graded Naming Test (/30)	4 (<5th)	6 (<5th)	26.0 (2.0)
Concrete synonyms (/25)	17 (10th)^{b}	21 (50th)^{b}	24.5 (0.7)
Abstract synonyms (/25)	12 (1-5th)^{b}	24 (75-90th)^{b}	24.3 (0.8)
Object Decision Task (/20)	19 (75–90th)	18 (50-75th)	17.8 (1.9)
DKEFS Design Fluency Task: switching	1 (<5th)	6 (50-75th)	n/d
Identification of unique visual entities			**n=17**
Famous Faces Test: Naming (/12)	4 (5th)	1 (<5th)	9.9 (1.7)
Famous Faces Test: Recognition (/12)	8 (10-25th)	1(<5th)	10.8 (1.2)
Landmark Naming (/15)	7 (<5th)^{a}	6 (<5th)^{a}	13.6 (1.7)
Landmark Recognition (/15)	8 (5th)^{a}	8 (5th)^{a}	13.7 (1.4)

Percentiles calculated from standardised tests, except where marked: a, calculated from previous healthy control sample (n=143); b, test administered with both visual and auditory presentation of words whereas the standardised percentiles are calculated for auditory presentation only; n/d = test not performed

Both QR and KL had evidence of anomia on the GNT, and QR had evidence of additional impairments of single word comprehension (abstract synonyms), surface dyslexia on the NART and executive function on design fluency; neither patient showed a deficit of short term memory or early visual perceptual function. On the standard Famous Faces Test (Warrington *et al.*1967) QR performed at the 5th percentile on face naming and normally on the face recognition component of the test, while KL showed impairments on both tasks. On a test assessing naming and recognition of 15 famous London landmarks from photographs, QR and KL each performed below the 5th percentile for naming and at the 5th percentile for recognition.

Subjects' exposure to the media.

The background media exposure of the control subjects and both patients was assessed formally as a potentially relevant factor influencing voice recognition ability. All subjects completed a questionnaire asking them to estimate their average media exposure for the preceding three months according to one of six categories: number of hours per week spent watching television, number of hours per week spent listening to the radio, and number of times per week they watched or read the news. QR's media exposure was greater than the average experience of the controls in each of these categories: on average each week she spent over 20 hours watching television (median control category=5-10 hours per week, range: 0-20+), over 20 hours listening to the radio (median control category=5-10 hours per week, range: 0-20+), and read or watched the news more than 10 times (median control category=8-10 times per week, range: 0-10+). In contrast control KL reported lower than average exposure in all categories but fell within the control range. He currently listened to the radio 0-1 hours per week (a change from 1-5 hours five years previously), he read the news once a week, and did not regularly watch television.

3.2.2. Experimental investigations

Semantic tests of voice, face and name recognition

Voice recognition was assessed using tests of familiarity, naming, identification and cross-modal matching of famous voices, and was compared to performance on parallel tests of face processing to examine the modality specificity of any deficits. In addition familiarity for written and spoken names was compared to performance in the voice modality. Methods are described in more detail in Chapter 2 (Section 2.6).

Comparison with identification of lower frequency faces

As the majority of public figures selected were recognised better from face than from voice by controls, and face recognition performance may be primed by previous presentation of the corresponding voices, therefore an alternative set of 24 difficulty-matched faces (see list of public figures in Appendix A.3) were selected from the pilot study stimuli which matched the set of famous voices in terms of accuracy of naming and identification. Identification achieved by 77% of pilot study group controls; mean=76.7, SD=8.7, range: 58-85%. Recognition by pilot study group controls was not significantly different between this set of faces and the 24 voices (Wilcoxon Ranksum Test: $z=-1.1$ $p>0.26$). This alternative set of faces was administered to QR and KL only.

Perceptual analysis of voices and faces

In order to assess the effects of perceptual factors on any voice recognition deficit, tests of analysis of low level vocal properties (speaker size and gender) and a test of apperceptive vocal processing: "difficult" speaker discrimination were performed. Perceptual processing of faces was also assessed using the Benton facial recognition test.

Recognition of vocal emotions

Processing of vocal emotion by QR and KL was assessed using 40 nonverbal vocalisations, 10 representing each of the emotions happiness, sadness, anger and fear, selected from a previously developed set (Sauter *et al.*2010; Sauter & Scott 2007). Items most reliably recognised by young normal subjects based on these previous normative data were selected. The subject's task on each trial was to select the emotion label describing the target emotion in a four-alternative forced choice format.

Identification of environmental sounds

40 common environmental sounds representing a variety of sound sources, including elemental sounds (e.g. thunder), man-made objects (e.g. kettle whistling), and animal calls (e.g. cow mooing), were selected from on-line databases. Environmental sounds were identified either by sound source (e.g. cow or a tap), or a description of the sound (e.g. mooing or dripping water); relatively lenient criteria for recognition were used, in line with the criteria used for voice identification. In a cross-modal version of the test, the subject was presented with arrays of four names and pictures, and required to match each sound with the correct name-picture combination.

Identification of musical instruments

Subjects were asked to name and identify 20 different sequentially presented instruments from their sounds (audio clips between 4 – 10 sec in duration), and then to identify the same instruments in a cross-modal matching condition, in which instrument sounds were presented together with arrays of four written instrument names and pictures. Cross-modal arrays contained the target instrument, a within-instrument family distractor (e.g. woodwind, brass, strings, percussion, and keyboard), and two instruments from a different instrument family. As QR had no musical training and KL had only two years of childhood piano lessons, patient performance was compared to 12 controls with up to two years musical training (defined as "inexperienced listeners": (Halpern *et al.*1995)).

3.3. Results

Table 3.2. Results of experimental tests assessing recognition of public figures from voice, face and name in patients and controls

	QR	KL	Case controls n=24	Control range
Voice			**mean (SD)**	**min-max**
Voice familiarity (/48) (% correct)	25 (52%)*	28 (58%)*	40.6 (4.0)	29-46
Voice naming (/24)	0*	0*	16.7 (4.4)	7-23
Voice identification (/24)	0*	0*	18.8 (3.9)	10-23
Cross-modal matching to face/name (/24)	3*	3*	23.5 (0.9)	21-24
Face				
Face familiarity (/48) (% correct)	29 (60%)*	31 (64%)*	46.7 (1.6)	43-48
Face naming (/24)	6*	3*	21.4 (2.7)	16-24
Face identification (/24)	17*	4*	23.6 (0.8)	21-24
Cross-modal matching to name (/24)	19[b]	11[b]	24.0 (0.0)	24-24
Difficulty matched faces: naming (/24)	6*	1*	14 (6.8)[a]	2-24
Difficulty matched faces: identification (/24)	13	1*	19 (5.6)[a]	3-24
Name				
Name familiarity (/48) (% correct)	43 (90%)	33 (69%)*	46.6 (1.6)	42-48

[a] Pilot control sample (n=26) scores for identification of 24 faces (see Appendix A.1) were used to assess performance on this test; [b]All control subjects performed at ceiling on this test, therefore single case statistics could not be performed; *patient scored significantly worse than control group ($p<0.05$)

3.3.1. Familiarity of voices, faces and personal names

Table 3.2 shows the results of familiarity judgments on voices, faces and names, in QR, KL and controls. For controls, the voice familiarity task was most difficult (mean score equivalent to 85% correct), compared to near-ceiling performance on face and name familiarity (mean score equivalent to 97% correct in each modality). QR performed close to chance (and significantly worse than controls: $t = -3.8$, $p<0.01$, $df=22$) for voice familiarity judgments; for face familiarity

judgments, QR's performance was above chance but also significantly worse than controls ($t = -10.8$, $p<0.001$, $df=22$), while for name familiarity judgments her performance was significantly worse than the control mean ($t=-2.2$, $p=0.04$, $df=22$) but within the control range. Further analysis of errors made by QR revealed that she correctly classified only 15/24 familiar voices as familiar, and misclassified 14/24 unfamiliar voices as familiar. On name familiarity she correctly classified 19/24 familiar names as familiar and misclassified 0/24, while she correctly classified 19/24 familiar faces as familiar, but misclassified 14/24 unfamiliar faces as familiar (i.e., she showed an inflated false alarm rate, especially for face familiarity: 14/19 errors). KL's performance was significantly worse than controls for all three modalities (voices: $t = -3.1$, $p<0.01$, $df=22$, faces: $t = -9.6$, $p<0.001$, $df=22$, names: $t = -8.3$, $p<0.001$, $df=22$). Analysis of KL's errors revealed a hit rate of only 6/24 familiar voices, 11/24 familiar faces and 14/24 familiar names. He made few false alarms: only 2/24 unfamiliar voices, 4/24 unfamiliar faces and 5/24 unfamiliar names were classed as familiar.

The difference between familiarity judgement performance to famous voices compared to their corresponding faces or names in case QR did not reach statistical significance (faces: $\chi^2 =0.90$, $p=0.34$, $df=1$, names: $\chi^2 =1.46$, $p=0.23$, $df= 1$). In case KL, the difference between familiarity judgement performance to famous voices compared to their corresponding faces did not reach statistical significance, (faces: $\chi^2 =1.46$, $p=0.23$, $df= 1$), however voice performance was found to be significantly different to performance on the corresponding names (names: $\chi^2 =4.90$ $p<0.05$, $df=1$). The difference between familiarity judgement performance to famous faces compared to their corresponding names was not significant in either patient (QR: $\chi^2 =0.25$, $p=0.62$, $df= 1$; KL: $\chi^2 =0.57$, $p=0.45$, $df= 1$).

3.3.2. Naming, identification and cross-modal matching of voices and faces

Table 3.2 shows the results of identification tasks for voices and faces in QR, KL and controls. Controls performed significantly better on tests assessing identification of faces than voices (naming: $t=5.9$, $p<0.001$, $df=21$; identification: $t=6.1$, $p<0.001$, $df=21$); face identification test performance was near ceiling. Both QR and KL performed at floor and significantly worse than controls for both naming ($t = -3.7$, $p<0.01$, $df=21$) and identification ($t = -4.7$, $p<0.001$, $df=21$) of famous voices, therefore the difference between modalities was not computed. Both patients performed significantly worse than controls for face naming (QR: $t = -5.6$, $p<0.001$, $df=22$, KL: $t = -6.7$, $p<0.001$, $df=22$) and face identification (QR: $t = -8.1$, $p<0.001$, $df=22$, KL: $t = -24.0$, $p<0.001$, $df=22$), however QR's performance improved substantially for identification of faces

compared with voices, and her performance was significantly superior to KL's (χ^2 = 14.31 p<0.001, df=1).

On cross-modal matching tasks, control performance was near ceiling for both voices and faces. For cross-modal matching of faces to names, both QR and KL performed worse than controls: all control subjects performed at ceiling on this test and single case statistics could not be performed, however QR's performance was significantly better than KL's (χ^2 = 5.89 p<0.05, df=1). For cross-modal matching of voices to faces and names, both patients performed at chance and significantly worse than controls (t = -22.2 p<0.001, df=19). The difference between cross-modal matching performance to famous voices compared to their corresponding faces was highly statistically significant for case QR (χ^2 =12.50, p<0.001, df= 1), but non-significant in case KL (χ^2 =3.5, p=0.06, df=1).

3.3.3. Identification of lower frequency faces

On identification of difficulty matched faces (Table 3.2), QR's performance did not differ significantly from healthy controls for either face naming (t = -1.2, p= 0.26, df=24) or identification (t=-1.1, p=0.30, df=24). KL's performance remained significantly inferior to controls (naming: t = -2.8, p<0.05, df=24; identification: t = -3.2, p<0.001, df=24).

Comment

The experimental control group here had a high average NART IQ (120.9, SD=6.3) and a greater mean number of years of education than QR, raising the possibility that a generic factor such as IQ contributed to her voice recognition deficit. A premorbid estimate of QR's IQ was not available, and any estimation based, for example, on demographic factors such as occupation would need to be made with caution in the individual case. Moreover, regression analysis in a larger control sample of older adults (n=48) (pilot study described in Appendix A.1), showed no evidence of association between number of years of education or NART IQ and voice recognition performance. In order to further explore any IQ-related contribution to QR's voice recognition difficulty, her performance on the voice recognition tasks was compared with five healthy control subjects (three female, two male) who had an average IQ typical for the greater London population (mean IQ 107.6, SD 6.7, range 96-112). This control group included three controls from the experimental control group with lower IQs (mean IQ 110.3, SD 2.1, range: 108-112) and two additional older adult controls (IQs 96 and 111) not included in the main study as they did not complete the perceptual voice tests. QR's performance was significantly inferior to this

lower-IQ control subgroup ($p<0.001$) on the voice familiarity ($t = -6.7$, $p<0.001$, $df=4$), naming ($t = -5.0$, $p<0.001$, $df=4$) and identification ($t = -6.0$, $p<0.001$, $df=4$) tasks.

Summary

These findings support the concept of a relatively modality-specific deficit of voice recognition in QR, in contrast to the multimodal deficit of person recognition exhibited by KL. At familiarity judgements QR performed close to chance for voices; in addition, her ability to judge the familiarity of faces was also clearly impaired, whereas her ability to judge the familiarity of names was somewhat less impaired. KL performed similarly whether judging the familiarity of public figures from voice, face or name, supporting a multimodal person familiarity deficit. Impairments of voice recognition were evident across the identification and cross-modal matching procedures used here in patient QR. Her ability to retrieve proper names from voice or face was clearly impaired, as anticipated on the basis of her general word retrieval impairment (Table 3.1). However, her ability to identify the same public figures from face information in the identification and cross-modal matching conditions (which do not rely on naming), though deficient to healthy controls, was clearly superior to her ability to identify voices, and superior to KL's performance in either modality. QR's score on the voice identification task was also highly significantly worse than the lower IQ control group: it therefore seems unlikely that her voice recognition deficit was due to IQ factors.

In line with previous work, control voice recognition scores were significantly lower than face recognition scores (Hanley *et al.*1998). In the pilot control regression analysis (see Appendix A.1), increased news exposure was positively associated with voice identification score. It is unlikely this factor explains QR's voice recognition deficit, as QR rated in the highest category for the number of times per week she read or watched the news (see subject details in Section 3.2.1). QR's relatively good performance on face identification appears initially somewhat paradoxical in relation to her poor performance on the face familiarity judgment: however, this pattern is likely to reflect an inflated false alarm rate (14/19 errors) on the face familiarity task.

Table 3.3. Results of experimental tests of perceptual processing of voices and faces in patients and controls

	QR	KL	Controls n=21	Control range
Voice perception			**mean (SD)**	**min-max**
Gender discrimination (/24)	24	24	n/a	n/a
Size discrimination (/32)	29	25	28.8 (4.7)	17-32
Speaker discrimination (/48)	39	33	35.0 (3.1)	29-41
Face perception				
Benton Facial Recognition Test (/56)	48	41	n/a	n/a

n/a = test not performed

3.3.4. Perceptual analysis of voices and faces

Table 3.3 shows the results of perceptual analysis tasks for voices and faces in QR, KL and controls. Both patients were able to judge gender and speaker size, and their performance was not significantly different to healthy controls (QR: $t = 0.0$, $p=0.97$, $df=20$; KL: $t = -0.8$, $p=0.44$,$df=20$). On the speaker discrimination task, QR's performance did not differ significantly from controls ($t = 1.3$, $p=0.22$, $df=20$) (Table 3.3). KL's performance was also not significantly different from controls (sample: $t = -0.6$, $p=0.54$, $df=20$). Both QR and KL performed normally on the Benton test of perceptual matching of faces.

Summary

This provides evidence that pre-semantic vocal processing mechanisms were intact in QR and KL. An impaired ability to identify voices was unlikely to be grounded in an early vocal perceptual deficit in either patient, and both patients were able to achieve an intact representation of individual voices as auditory objects sufficient to discriminate between different speakers, yet were unable to gain a sense of familiarity to a voice or to associate these representations with other stored information about familiar speakers.

Table 3.4. Results of experimental tests of recognition of vocal emotions, environmental sounds and musical instruments in patients and controls

	QR	KL	Controls mean (SD) n=12	Control range min-max
Vocal emotion recognition				
Cross-modal matching to emotion (/40)	32	30	35.1 (3.1)[c]	26-39
Environmental sounds				
Environmental sound identification (/40)	35	34	37.1 (2.1)[a]	33-39
Cross-modal matching to picture/name (/40)	39	40	39.9 (0.3)[b]	39-40
Musical instruments				
Instrument sound name (/20)	5*	6*	13.1 (2.8)	8-18
Instrument sound identification (/20)	6*	7*	13.7 (2.9)	9-18
Instrument picture name (/20)	4*	11*	17.1 (1.7)	14-19
Instrument picture identification (/20)	10*	13*	17.3 (1.5)	15-19
Cross-modal matching sound to picture/name (/20)	12*	18	19.3 (0.8)	18-20

[a] n=14 controls; [b]n=10 controls; [c]Separate control group results (n=22), were used to assess performance on this test, *patient scored significantly worse than control group ($p<0.05$)

3.3.5. Recognition of vocal emotions

Table 3.4 shows the results of the vocal emotion recognition test for QR, KL and controls. Both QR and KL performed comparably to healthy controls (QR: $t = -1.0$, $p=0.34$, $df=21$, KL: $t = -1.6$, $p=0.12$, $df=21$).

3.3.6. Identification of environmental sounds

Table 3.4 shows the results of environmental sounds identification tests for QR, KL and controls. On the sound identification test, both QR and KL performed comparably to healthy controls (QR: $t = -1.0$, $p=0.35$, $df=14$; KL: $t = -1.4$, $p=0.18$, $df=14$). On the cross-modal matching task, KL performed at ceiling and QR near ceiling; 9/10 control subjects performed at ceiling on this task.

3.3.7. Identification of musical instruments

Table 3.4 shows the results of musical instrument identification tests for QR, KL and controls. Inexperienced listeners recognised on average 68.5% (SD=14.4%) of the instruments, an accuracy level inferior to identification of famous voices by the same controls. Both patients performed significantly worse than controls on tests of instrument sound naming (QR: $t = -2.8$,

$p<0.05$, df=11; KL: $t = -2.4$, $p<0.05$, df=11) and identification (QR: -2.6, $p<0.05$, df=11; KL: $t = -2.2$, $p<0.05$, df=11). On the cross modal matching task QR performed above chance, however her score was significantly different to controls ($t = -9.4$, $p<0.001$, df=11); in contrast KL's performance was not significantly different to controls ($t = -1.7$, $p=0.12$, df=11). Both controls' and patients' performance improved on the visual version of the task. Both patients' scores were significantly different to controls on tests of instrument picture naming (QR: $t = -7.4$, $p<0.001$, df=11; KL: $t = -3.4$, $p<0.01$, df=11) and identification (QR: $t = -4.7$, $p<0.01$, df=11; KL: $t = -2.8$, $p<0.05$, df=11).

Summary

Both QR and KL performed essentially normally on tests of environmental sound and vocal emotion recognition. These findings suggest that the deficit of voice recognition exhibited by each patient is at least relatively specific for human voices, and supports dissociation between vocal identity and emotion processing in these patients. QR's ability to recognise another category of finely differentiated sounds (musical instruments) was impaired, though superior to her ability to recognise voices. In contrast, KL exhibited normal auditory recognition of instruments on the cross-modal matching task. This pattern of results might signify that QR has an auditory agnosia that affects recognition of voices and certain other categories of auditory objects, whereas KL has a primary deficit of person knowledge. However, this interpretation requires some qualification, since both patients also exhibited impaired visual recognition of instruments relative to the healthy control group, while QR scored lower on both the auditory and pictorial versions of the task relative to KL. It is difficult to equate musical exposure between non-musicians (KL's musical experience is likely to have been wider than QR's) and this may also be affected by other factors, such as general educational attainment (QR had fewer years of formal education than KL). These factors are likely a priori to be relatively more important for music than person knowledge. Moreover, no other category of complex nonverbal sounds is truly comparable in diversity and familiarity to human voices (for practical purposes, a musically untrained subject is likely to be acquainted with perhaps ten musical instruments, but potentially hundreds of individual human voices).

3.4. Discussion

In this study neuropsychological evidence is presented for distinctive deficits of voice recognition in two patients with focal neurodegenerative disorders. The first patient, QR, exhibited severe impairments of voice identification and familiarity judgments with preserved identification of

difficulty-matched faces and environmental sounds; recognition of another highly differentiated category of complex sounds (musical instruments) was better than recognition of voices albeit impaired. In contrast, patient KL exhibited severe impairments of both voice and face recognition, partly preserved recognition of musical instruments and essentially normal identification of environmental sounds. Both patients demonstrated preserved ability to analyse perceptual properties of voices to the level of individual speaker discrimination and to recognise emotions in voices. The profiles of deficits exhibited by both QR and KL are summarised in Table 3.5. QR's deficit of voice processing could be characterised as a failure to associate familiar voices with other specific semantic information about the individual: associative phonagnosia. Further, this deficit is relatively selective for voices. KL's more uniform deficit of recognition across modalities (voices, faces and names) suggests a multimodal failure of person knowledge with associative phonagnosia as one component.

Table 3.5. Summary of experimental neuropsychological profiles in QR and KL

	Domain	**Case QR**	**Case KL**
Voices	Identification	↓	↓
	Familiarity	↓	↓
	Emotion recognition	**N**	**N**
	Perception	**N**	**N**
Other sounds	Musical instrument recognition	↓↓	↓
	Environmental sound identification	**N**	**N**
Faces	Recognition	**N**†	↓↓
	Perception	**N**	**N**

N normal performance, ↓ impaired performance relative to controls, ↓↓ impaired performance relative to both controls and other case; †when matched to voices for difficulty

Detailed studies of phonagnosia are comparatively few (Garrido *et al.*2009; Neuner *et al.*2000; Van Lancker *et al.*1987; Van Lancker *et al.*1982; Van Lancker *et al.*1988; Van Lancker *et al.*1989) and neuropsychological investigations of voice recognition have generally been undertaken in patients presenting with acquired or developmental prosopagnosia (Gainotti *et al.*2008; Gentileschi *et al.*1999; Gentileschi *et al.*2001; von Kriegstein *et al.*2006). Selective phonagnosia has recently been described on a developmental basis (Garrido *et al.*2009): this individual had deficits of voice recognition and familiarity despite normal face recognition.

Deficits in person knowledge are well described as a presentation of right temporal lobe degeneration: selective impairment of face recognition and multimodal impairment extending to recognition of voices and names have been described. However, phonagnosia has not previously been emphasised as the leading feature of person knowledge breakdown in degenerative disease, and detailed anatomical correlates of this deficit remain to be established.

Modality-specific deficits of person knowledge (in judgements of familiarity and also retrieval of semantic information) present a potentially critical test of the IAC model of person recognition (described in Section 1.7.1), and indeed models of the semantic system more broadly (Gainotti 2007a; Lucchelli *et al.*2008; Mahon, Anzellotti, Schwarzbach *et al.* 2009; Snowden *et al.*2004; Thompson *et al.*2004). The multimodal impairments displayed by KL here (and by most previously studied patients with progressive prosopagnosia) are consistent with a core defect affecting a multimodal store of knowledge about familiar people (Gainotti 2007a; Gainotti *et al.*2008; Gentileschi *et al.*2001; Lucchelli *et al.*2008), reflecting either damage to the stores proper or a disconnection from the PIN. However, QR exhibits a relatively selective associative deficit of voice recognition.

Such a deficit could in principle arise at pre-semantic stages in the voice processing pathway: the demonstration of intact early vocal perceptual analysis and speaker discrimination in QR would be consistent with a dissociation of perceptual descriptions or voice recognition units from the PIN. A lesion at this processing stage might also account for loss of the sense of familiarity of voices. However, while voices are often analogised as 'auditory faces', the demands of perceptual analysis differ substantially between the auditory and visual modalities, and mechanisms for the perceptual analysis of voices remain poorly understood. Deriving a faithful neural representation of a voice is likely to depend on intact mechanisms for processing timbre (Griffiths *et al.*2004; Warren *et al.*2006). Selectivity of voice recognition deficits could arise from an abnormal interaction between combinations of complex vocal properties such as timbre, articulation, prosody which distinguish an individual's voice (Perrachione & Wong 2007; Remez *et al.*1997; Schweinberger 2001; Van Lancker *et al.*1985), and subsequent stages of voice identity processing (it is of interest that KL reported some difficulty understanding unfamiliar accents). Interaction between perceptual and semantic mechanisms of voice processing would be in line with recent re-evaluations of models of person identification (Lucchelli *et al.*2008), and may be particularly critical under non-standard listening conditions (e.g., identification of voices over the phone or when singing: (Benzagmout *et al.*2008; Garrido *et al.*2009)).

A related issue is the specificity of agnosia for voices versus other kinds of complex sounds and versus unique entities (i.e., items associated with proper nouns: (Gainotti *et al.*2008)) in sound or other modalities. This speaks to the more fundamental issue of the degree of specialisation of brain mechanisms for processing voices versus other kinds of ecologically relevant complex sounds (Belin *et al.*2004). Both QR and KL were able to recognise environmental sounds successfully, arguing against a generalised auditory agnosia: this dissociation corroborates previous findings (Garrido *et al.*2009; Neuner *et al.*2000; Peretz *et al.*1994). QR and KL demonstrated comparably weak performance for recognition of London landmarks, but in both cases this was clearly superior to recognition of voices (and in the case of KL, also superior to recognition of faces). Furthermore, QR demonstrated a clear superiority for recognition of faces versus voices. Taken together, these observations argue that phonagnosia in these cases is unlikely simply to reflect a generic defect of fine-grained semantic attributions.

Within the auditory modality, both QR and KL showed superior recognition of musical instruments compared with voices, however QR's performance was clearly inferior both to healthy controls and KL. Musical instruments are themselves likely to constitute a specialised category of complex sounds, but (unlike voices) cannot strictly be considered 'unique entities': nevertheless, the pattern of QR's results raises the possibility that her phonagnosia is part of a broader defect of differentiation amongst closely related auditory entities, which could in turn arise at the level of associative (semantic) processing or as a result of an abnormal interaction between perceptual and semantic mechanisms. This formulation would be consistent with evidence in the visual domain, in both the present and previous studies (Gainotti 2007a; Gainotti *et al.*2008): patients with right temporal lobe lesions in general exhibit a more severe deficit for recognition of faces than landmarks and other unique visual entities, however this recognition deficit is seldom restricted purely to faces.

The present study shares the limitations of single neuropsychological case studies, including limited scope for anatomical correlation: this applies particularly to neurodegenerative pathologies, in which any regional selectivity of brain damage is relative rather than absolute. That caveat aside, these cases together illustrate a syndrome of progressive associative phonagnosia and demonstrate that this may be relatively selective with respect to other stages of voice analysis, other aspects of person knowledge and other categories of auditory objects. Important directions for future work include the longitudinal study of the evolution of

phonagnosia in relation to other defects of person knowledge in patients with degenerative pathologies, a more detailed examination of the processing of other unique or highly differentiated auditory entities in phonagnosic individuals, and structural and functional anatomical substrates for the syndrome.

4. Study 2: A neuropsychological and neuroanatomical analysis of voice processing in tvFTLD and AD

4.1. Introduction

The majority of reported cases of phonagnosia in degenerative diseases propose that voice recognition becomes affected with evolution of the progressive prosopagnosia syndrome (Evans *et al.*1995; Gainotti 2007a; Gainotti *et al.*2008; Gentileschi *et al.*1999; Gentileschi *et al.*2001; Joubert *et al.*2006) whereas selective phonagnosia has seldom been reported. This may be because voice recognition has rarely been tested in cases without prosopagnosic symptoms or because phonagnosia is less clinically salient. In Study 1 (Chapter 3), associative phonagnosia was described both in a case of bvFTD with relatively-preserved face and proper name recognition and in a patient with right tvFTLD who showed multimodal deficits of person recognition; however with only two subjects in this study there was limited scope either for neuroanatomical correlation of deficits or for assessing how widely selective phonagnosia presents in FTLD. The second study in this thesis aimed to investigate the neuropsychological and neuroanatomical signatures of voice processing in two patient groups with FTLD and AD, dementias with pathology affecting regions in the temporal lobes. In targeting these disease groups it was recognised that FTLD is clinically and anatomically heterogeneous, whereas AD typically presents with a more uniform clinical and anatomical profile. In particular, the subgroup of patients with FTLD who have predominant temporal lobe atrophy (the subgroup predicted a priori to develop voice processing deficits) have heterogeneous clinical presentations, including both SD (progressive semantic aphasia or progressive prosopagnosia) and progressive behavioural decline (bvFTD), particularly if atrophy chiefly affects the right temporal lobe (Chan *et al.*2009). Accordingly, for the purposes of the present 'lesion-led' study an anatomical criterion was used for selecting patients with tvFTLD based on the presence of predominant temporal lobe atrophy: this noncanonical, anatomically defined subgroup was termed 'temporal lobe variant frontotemporal lobar degeneration (tvFTLD)' (Brambati *et al.*2009).

Study 1 demonstrated that associative phonagnosia may occur independent of deficits of performance on perceptual tests, in line with current cognitive models and previous neuropsychological evidence concerning the organisation of voice processing (see Section 1.4.2) (Belin *et al.*2004; Ellis *et al.*1997; Hanley *et al.*2009; Lucchelli *et al.*2008; Neuner *et al.*2000; Schweinberger *et al.*1997; Van Lancker *et al.*1988) as well as frequent descriptions of selective impairments to multimodal semantic knowledge in FTLD affecting the ATLs. Semantic deficits

of person knowledge (recognition of faces and names) have also been described in AD (Greene *et al.*1996), but voice recognition has rarely been tested. Impairments of high level auditory processes such as recognition of linguistic prosody in AD have been primarily attributed to posterior cortical dysfunction (Allender *et al.*1989; Roberts *et al.*1996; Taler *et al.*2008; Testa *et al.*2001), therefore an apperceptive impairment of voice processing may be relatively more prominent in this patient group. As in Study 1, tests of early encoding, discrimination as well as recognition of voices were performed in the target clinical groups. As models of auditory object processing suggest that processing categorical information about the characteristics of complex sounds may depend on the interaction between perceptual encoding mechanisms and 'top-down' semantic factors the relations between perceptual and semantic task performance were analysed in each patient group.

Temporo-parietal atrophy in AD has been associated with a particular pattern of neuropsychological deficits, including impairments on naming, working memory and arithmetic tasks (Stopford *et al.*2008). The relationship between voice processing and other cognitive tasks is also of interest to understanding variability in neuropsychological profiles of each disease group. In tvFTLD the relations between voice recognition performance and performance on other semantic processing tasks is of interest to models of semantic processing which debate whether there is segregation of category or modality-specific information in the ATLs. In both syndromes, the relations between voice measures and disease severity measures may help to understand how phonagnosic and multimodal person recognition symptoms relate to the progression of the diseases. Atrophy affecting frontal, temporal and parietal cortices in these diseases allows us to identify critical nodes in functional and anatomical cerebral networks delineated in functional imaging studies of voice processing.

The neuroanatomical correlates of voice processing performance were assessed using VBM (Ashburner *et al.*2000). The results were considered in relation to the anatomical hierarchy hypothesized from perceptual analysis of voices in posterior temporal cortices to associative processing of voices and other modalities of person knowledge in more ATL areas (Belin *et al.*2002; Bishop *et al.*2009; Imaizumi *et al.*1997; Nakamura *et al.*2001; Van Lancker *et al.*1988; Van Lancker *et al.*1989; von Kriegstein *et al.*2006; Warren *et al.*2006). Voice processing was assessed in relation to face and name processing, as the cognitive and neural architecture of voice processing and its relations to other modalities of person knowledge continue to be defined.

On the bases of the patterns of atrophy in the two diseases, distinct profiles of phonagnosia were hypothesised in tvFTLD and AD, with more severe associative impairment in tvFTLD and relatively more prominent apperceptive impairment in AD. Based on anatomical evidence in the healthy brain it was hypothesised that semantic deficits in processing voices (in common with other kinds of person knowledge) would be associated with atrophy of ATLs: particularly in anterior STS/STG (Olson *et al.*2007), and voice apperceptive deficits would be associated with atrophy of posterior temporo-parietal regions.

4.2. Materials and methods

4.2.1. Subject demographic characteristics and clinical details

14 consecutive patients with tvFTLD, 22 patients with AD and 35 healthy older control subjects participated. Demographic characteristics are summarised in Table 4.1. Subject groups did not differ significantly in age, gender distribution, or years of education (all $p>0.05$); the tvFTLD and AD groups did not differ significantly on two general measures of clinical severity (symptom duration and Mini Mental State Examination (MMSE) score).

Patients with tvFTLD were selected based on the presence of selective, bilateral ATL 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 tvFTLD had a syndrome of SD according to the consensus criteria of Neary et al. (Neary *et al.*1998); within this SD subgroup, 10 patients presented with progressive semantic aphasia and three presented with progressive prosopagnosia. One patient within the FTLD group presented with bvFTD. 20/22 patients with a clinical diagnosis of AD had brain MRI: 16 patients had disproportionate symmetrical hippocampal atrophy and four had generalised cerebral atrophy.

Table 4.1. Summary of subject characteristics

	tvFTLD n=14		**AD n=22**		**Healthy controls n=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 (yrs)	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	SD (n = 13)[a] bvFTD (n = 1)		Amnestic AD		n/a	
Symptom duration (yrs)	5.4 (1.7)	3-8	5.7 (2.4)	2-11	n/a	
MMSE (/30)	21.1 (7.2)**	6-29	21.3 (4.2)**	14-28	29.4 (0.6)[b]	28-30
Medication	4[c]		18[d]		n/a	
Cardinal symptoms						
Voice recognition	n = 9 (64%)		n = 11 (50%)		n/a	
Face recognition	n = 7 (50%)		n = 8 (36%)		n/a	
Voice familiarity	n = 3 (21%)		n = 3 (14%)		n/a	
Face familiarity	n = 5 (36%)		n = 2 (9%)		n/a	
Media exposure						
TV watching (hrs per week)	15.1 (9.2)	0-32	15.9 (10.0)	0-35	14.4 (10.5)	0-63
Radio listening (hrs 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

n/a 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 ($p<0.01$); [a] 10 cases with progressive semantic aphasia, 3 cases with progressive prosopagnosia; [b] 23 controls performed MMSE; [c]two patients taking a serotonin reuptake inhibitor, one taking anti-Parkinson's medication, one taking lithium; [d]16 patients taking a cholinesterase inhibitor, 2 taking memantine

Clinical details of patients

Within the tvFTLD group, 10 of the 13 patients with a diagnosis of SD presented primarily with difficulties with language: word finding, word comprehension and naming; 3 of these patients had additional symptoms at presentation, including episodic memory complaints and changes in personality (withdrawal and apathy). The three remaining patients in the SD subgroup presented primarily with difficulties with face recognition but not word finding difficulties. One patient with predominantly right sided temporal lobe atrophy presented with changes in personality (apathy), episodic memory, impaired attention, increased religiosity, auditory hallucinations, and decline in visual memory on neuropsychological testing. This patient was taking lithium for depression, diagnosed 6 years prior to the assessment. Two other patients with tvFTLD were

taking a serotonin reuptake inhibitor for mood symptoms; one patient with tvFTLD was on treatment for Parkinsonism.

All patients in the AD group had episodic memory difficulty as a leading symptom; 5 patients had additional symptoms at presentation including changes in mood or difficulties with topographical memory or reading. When assessed 16 patients were taking an acetylcholinesterase inhibitor (donepezil or galantamine), and two patients were taking memantine. Two patients with AD were unable to have MRI due to a cardiac pacemaker; computed tomography in one of these patients showed generalised cerebral atrophy.

Information about background media exposure for all subjects was obtained (see Methods Section 2.2.3 for further details), in which control subjects and patients' carers estimated average number of hours spent each week watching television and listening to the radio and the average number of news exposures each week (over the period of previous three months). Results are summarised in Table 4.1. The AD group did not significantly differ from the healthy control group in any category. The tvFTLD group did not differ significantly from the AD or healthy control groups for estimated television exposure but had significantly lower ($p < 0.05$) estimated radio and news exposure than both the control group and the AD group.

Face and voice recognition symptoms

Although patients were not selected for inclusion in the study based on a history of phonagnosia, the frequency of voice and face recognition difficulties in the target disease groups was of interest. Patients' carers were given a brief questionnaire asking them to report if the patient had any difficulty identifying the voices of people they should know well (e.g., over the telephone); and if so, whether the patient showed that they were familiar with the voice. Analogous questions were posed for face recognition. 9/14 tvFTLD patients and 11/22 AD patients were reported to have some difficulty with recognition of voices; 7/14 tvFTLD patients and 8/22 AD patients were reported to have some difficulty with recognition of faces. Loss of voice familiarity was reported in 3/14 tvFTLD patients and 3/22 AD patients, while loss of face familiarity was reported in 5/14 tvFTLD patients and 2/22 AD patients.

Assessment of peripheral hearing

The audiometry procedure is described in Chapter 2 (Section 2.2.2). Most subjects had no clinical history of hearing loss. The mean value of response time (i.e., detection threshold) in the right ear

for each frequency was taken as the detection threshold for that frequency except in four subjects (1 tvFTLD, 1 AD, 2 controls) that reported unilateral right-sided hearing loss and in these subjects the left ear was tested. One control subject had mild bilateral high frequency hearing loss, previously confirmed on clinical audiometry. One tvFTLD patient had otosclerosis corrected with hearing aids and another had mild cochlear-type (mid-frequency) hearing loss of genetic origin confirmed on clinical audiometry. One AD patient had mild bilateral high frequency hearing loss and another AD patient had post-infectious unilateral hearing loss, previously confirmed on clinical audiometry. 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. As anticipated, increasing age was associated with a significant increase in mean response time (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: tvFTLD group= 7.2dB, AD group= 4.7dB); these threshold elevations were small and unlikely to be clinically relevant. There were no significant differences between either patient group and the control group at any other frequency tested, and no significant differences were observed between tvFTLD and AD groups at any frequency.

4.2.2. Subject background neuropsychological assessment

Results are presented in Table 4.2. Relative to the healthy control group both the tvFTLD group and the AD 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 AD group showed additional deficits of visual object perception and auditory verbal working memory. The tvFTLD group had lower verbal and reading IQ than the AD group and performed significantly worse than the AD group on tests of naming, tests of semantic knowledge (verbal comprehension and London landmark recognition tests), reading and recognition memory for faces; while the AD group performed significantly worse than the tvFTLD group on tests of nonverbal reasoning and recognition memory for words. Considered together, these patterns of performance support the clinical and neuroanatomical classification for each disease group.

Table 4.2. Results of general neuropsychological assessment

Test (max score)	tvFTLD n=14		AD 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[d]	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)**†[f]	7-24	20.9 (2.7) **[f]	13-24	24.3 (1.3)	19-25
Abstract synonyms (/25)	15.4 (5.8) **†[f]	8-24	20.9 (3.5) **[f]	14-24	24.3 (1.2)	20-25
Landmark name (/15)	2.6 (3.7) ** [g]	0-12	6.1 (4.0) **[g]	0-15	13.5 (1.3)[h]	11-15
Landmark recogn (/15)	4.6 (4.7) **†[g]	0-12	8.0 (4.1) **[g]	0-15	13.7 (1.2)[h]	11-15
Non-semantic skills						
GNT (/30)	2.2 (6.1)**‡	0-23	11.6 (7.9)**	0-26	26.0 (2.4)	19-30
Object decision task (/20)	16.5 (5.0)	8-29	15.8 (2.8)**	9-19	18.5 (1.2)	16-20
Digit span forward (/12)	7.3 (2.7)	4-12	7.1 (2.3)*	4-11	8.7 (2.0)	4-12
Digit span back (/12)	6.1 (3.3)	0-10	4.9 (2.7)*	0-10	7.4 (2.6)	2-12
Arithmetic (GDA) (/24)	8.9 (6.9)**	0-20	5.5 (4.5)**	0-14	15.4 (4.8)[c]	6-23
Episodic memory						
RMT words (/50)	35.4 (7.0)**	24-47	30.1 (7.3)**†	19-47	47.3 (1.8)[c]	43-49
RMT faces (/50)	28.9 (4.1)**‡	24-40	35.0 (5.6)**	25-45	42.2 (4.7)[c]	35-49
Executive function						
Stroop Word reading scaled	5.2 (4.0)[a]*	1-14	5.8 (4.6)**	1-13	10.7 (2.7)[c]	3-14
Stroop Inhibition scaled	6.3 (4.6)[b]**	1-13	3.6 (3.2)**	1-11	11.5 (2.0)[c]	7-14

*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 ($p<0.001$); WMS-R digit span forwards, backwards; GDA, Graded Difficulty Arithmetic; London landmark naming and recognition test; (Whiteley & Warrington 1978)Stroop, D-KEFS Stroop test scaled scores; WASI, Wechsler Abbreviated Scale of Intelligence; [a] 1 tvFTLD subject was unable to read the words and a scaled score of 1 was used. [b] n=12 (2 tvFTLD subjects were unable to name colours); [c] 19 / 35 controls tested on these tasks. [d] Reading IQ was measured on the NART unless the subject scored ≤15/50 on this test, in which case the Schonell Graded Word Reading Test IQ was used (Schonell & Goodacre 1971). [e] 1 tvFTLD subject did not perform recognition memory tasks. [f] 2 tvFTLD and 1 AD subject did not perform synonyms tests. [g] 3 tvFTLD and 2 AD subjects did not perform the London landmarks test. [h] 34 / 35 controls tested on these tasks.

4.2.3. Experimental tests

Experimental tests are described in detail in Chapter 2. Assessment of encoding of low-level voice attributes of gender and VTL, an index of vocal size were performed. Speaker discrimination was assessed using two versions of the test stimuli to create two levels of speaker discrimination task difficulty: a 'difficult' test in which inter-speaker variations in vocal pitch were fixed and an 'easy' test in which pitch was not fixed.

The Benton Facial Recognition Test (Benton *et al.*1989) was administered to assess unfamiliar face discrimination. Semantic tests assessing familiarity, identification, and cross-modal recognition of famous people in voice, face and name modalities were presented as described in Chapter 2 (Section 2.6).

4.2.4. Analyses of behavioural data

For experimental 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 tvFTLD and AD 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.

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, MMSE score was used for the AD group and British Picture Vocabulary Scale (BPVS; a measure of semantic impairment) for the tvFTLD group. Within the tvFTLD group, the general neuropsychological and experimental

test performance of subjects with predominantly left-sided versus predominantly right-sided temporal lobe atrophy was compared: differences in means are reported with 95% Wald type bootstrap confidence intervals (2000 replicates).

4.2.5. VBM analyses

For 18 AD subjects and 11 tvFTLD MRI images were acquired. Further description of pre-processing stages and analyses are described in Chapter 2 (Section 2.9).

For each modality in the experimental battery (voices, faces, names), associations between regional GMV and subtest performance were assessed in both disease groups using linear regression models. Where the interaction was found to be significant, the within-group associations were investigated further. In separate modality design matrices, GMV was modelled as a function of the experimental subtest score-by-group interaction term with group, age and TIV included as covariates. Where no significant group interaction was identified, GMV was modelled as a function of experimental subtest score in both disease groups, with covariates of group, age, and TIV. 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 matching 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.

Grey matter associations were assessed over the whole brain and within two regions of interest (see Section 2.9.2). A voxel-wise statistical threshold $p< 0.05$ FWE-corrected for multiple comparisons was applied in all analyses, a global $p<0.05$ FWE-corrected threshold was applied in the combined-modalities conjunction analysis. SPMs 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 performance in other modalities.

In addition to analyses for experimental tests, associations between grey matter volume and background semantic tests (BPVS, concrete and abstract synonyms tests, landmark naming and identification) were also investigated to compare with associations found for person recognition tasks. The same design and methodology as above to separate modality analyses was implemented, not all subjects performed the landmark recognition test: 17 AD subjects and 9 tvFTLD subjects were entered into the analyses.

4.3. Neuropsychological results

4.3.1. Perceptual analysis of voice attributes

Results for the patient and healthy control groups on early perceptual and apperceptive subtests for each modality are summarised in Table 4.3.

Table 4.3. Behavioural data: perceptual and apperceptive processing of voices and faces

Subtest (max score)	**tvFTLD n=14**		**AD n=22**		**Healthy controls n=35**		**tvFTLD – AD:**
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range	Difference in means (95% CI)
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† **(**-0.5, -0.01)
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)

CI, 95% confidence intervals; *significantly worse than controls (p< 0.05); **significantly worse than controls (P< 0.01%); †AD group significantly worse than the tvFTLD group (p < 0.05)

On the vocal gender subtest, the AD group performed significantly worse ($p<0.05$) than the healthy control group: this difference being driven by a subgroup of four AD patients (the remaining patients scoring at ceiling on this task); the performance of the tvFTLD group did not differ from healthy controls, however all subjects in tvFTLD 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 AD group performed significantly worse ($p<0.05$) than healthy controls. There were no significant performance differences between the tvFTLD group and healthy controls or between the two patient groups.

No correlations were seen between speaker discrimination and performance on the Benton face task in either patient group (see Appendix A.4.1).

4.3.2. Semantic analysis of voices

Recognition of voices, faces and names

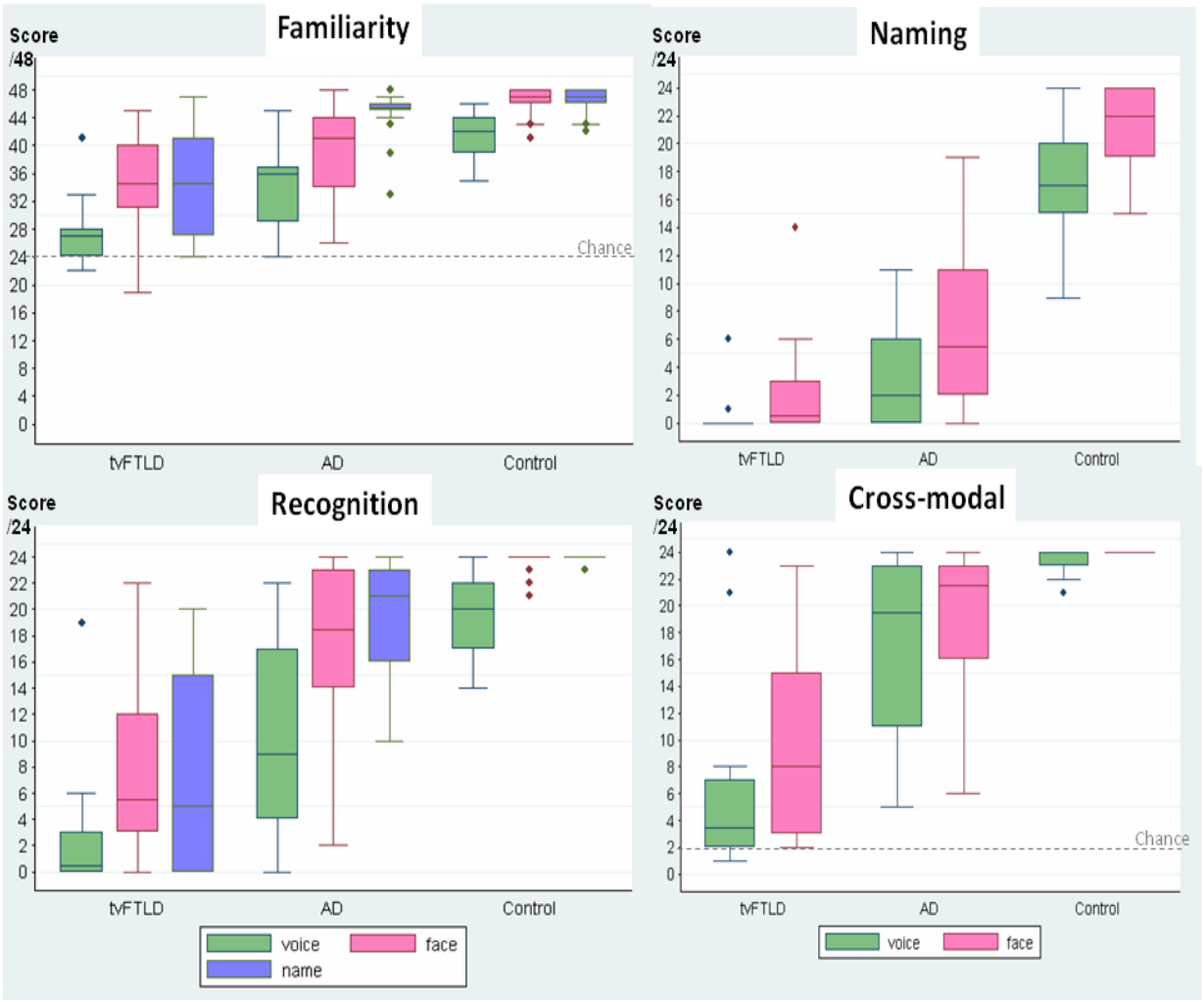
Results for the patient and healthy control groups on semantic subtests for each modality are summarised in Table 4.4.

Table 4.4. Behavioural data: semantic processing of voices, faces and names

Subtest (max score)	**tvFTLD n=14**		**AD n=22**		**Healthy controls n=35**		**tvFTLD – AD:**
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range	Difference in means (95% CI)
Familiarity							
Voice (/48)	27.5 (4.8)**	22-41	34.4 (5.5)**	24-45	41.5 (2.9)	35-46	-6.9‡ (-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‡ (-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† (-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† (-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‡ (-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‡ (-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‡ (-16.6, -8.2)
Cross-modal matching							
Voice (/24)	6.4 (7.1)**[a]	1-24	17.4 (6.4)**	5-24	23.6 (0.8)	21-24	-11.2‡ (-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‡ (-13.9, -5.0)

CI, 95% confidence intervals; * significantly worse than controls ($p<0.05$); ** significantly worse than controls ($p<0.001$); † tvFTLD group significantly worse than AD group ($p<0.01$); ‡ tvFTLD group significantly worse than AD group ($p<0.001$); [a] 1 tvFTLD patient 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 of 3/24 items.

Figure 4.1. Box plots to show tvFTLD, AD and control group semantic test scores



Boxes indicate median, 25th and 75th percentile values; whiskers indicate range of values.

On all semantic subtests, both the tvFTLD group ($p<0.001$) and the AD group ($p<0.05$) 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 tvFTLD group performed significantly worse than the AD 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 matching ($p<0.01$) and naming ($p<0.01$). The

tvFTLD group showed a significantly larger ($p<0.05$) decrease in score compared to the AD group for identification in the name modality compared with the voice modality; the tvFTLD-AD performance discrepancy did not differ significantly between modalities for the other subtests. In particular, there was no evidence that the magnitude of the tvFTLD-AD difference varied between voice and face modalities in any subtest.

Relationship between semantic subtest performances in disease groups

Within both the voice and face modalities, performance was significantly positively correlated ($p<0.05$) between all semantic subtests in the AD group, while in the tvFTLD group cross-modal matching was positively correlated with identification and familiarity (voices only). Examining correlations between modalities, identification and naming subtests were each significantly positively correlated ($p<0.05$) between voice and face modalities in both patient groups; while voice and face familiarity were positively correlated ($p<0.01$) in the AD group but not in the tvFTLD group (see Table A.4.2 in Appendices). Between the voice and name modalities, familiarity scores were not significantly correlated in either patient group, while identification scores were significantly correlated ($p<0.05$) in the AD group but not the tvFTLD group.

Relationship between perceptual and semantic task performance in disease groups

In the AD group, scores on the 'difficult' speaker discrimination test and voice familiarity were significantly correlated ($p<0.05$), whereas in the tvFTLD group, there were no significant correlations between vocal perceptual and voice recognition performance (see Appendix A.4.1).

4.3.3. The effect of disease severity

Disease severity is determined by the extent of pathological changes in the brain. In the absence of such markers the severity for both patient groups number of years of disease severity was utilised, on the basis of an informant's estimation of when symptoms were first noticed, and where that was not possible, estimation was taken from reviewing patient's clinical notes. A secondary measure of disease severity was examined in each patient group. MMSE score was utilised in AD subjects, a clinical measure of disease severity that is widely used in this disease. As the MMSE is a less useful measure in tvFTLD due to its over-reliance on verbal skills (Ridha & Rossor 2007), and due to core deficits in conceptual knowledge (in particular in SD), the British Picture Vocabulary Scale (BPVS) a general semantic task was used. An exploration of the association between disease severity and degree of voice processing impairment was tested for both measures.

Voice identification and familiarity performance were not associated with disease severity measures in either disease group (Table A.4.3 in Appendices). Only cross-modal recognition was positively associated with disease severity measures: in the tvFTLD group, it was positively associated ($p<0.01$) with BPVS score; while in the AD group, cross-modal voice recognition was signficantly associated ($p<0.05$) with MMSE score. No significant associations with disease severity measures were found for speaker discrimination or unfamiliar face discrimination in either group (see Appendix A.4.3).

4.3.4. The relationship of voice performance to other cognitive skills

In addition to the cognitive measures used as disease severity measures (MMSE and BPVS), the influence of background neuropsychological variables on voice task performance is of interest to understanding the mechanisms of and cognitive influences on voice processing. Within each patient group consideration was given to the relationship between voice task performance and background neuropsychological task performance in other potentially relevant cognitive domains, to facilitate understanding the pattern of cognitive impairments that may accompany deficits of voice recognition in each group. As vocal semantic sub-tests were impaired in both groups, the relationship between voice recognition performance and background neuropsychological and semantic tests was assessed. As vocal apperceptive deficits were observed in the AD group, the relationship between speaker discrimination tests and a sub-set of background neuropsychological tests directed to nonverbal processes and processes associated with temporo-parietal atrophy in AD (Stopford *et al.*2008) were assessed. These included nonverbal IQ, forwards and backwards digit span, naming on the GNT and tests of recognition memory and arithmetic.

Correlations of voice performance with neuropsychological measures

Correlations between voice semantic tests and background neuropsychological and semantic test scores in each patient group are displayed in Appendix A.4.4. In the tvFTLD group, there was limited evidence of correlation between semantic voice performance and performance on other semantic tests. Only voice naming positively correlated with performance on the BPVS and concrete synonyms, cross-modal matching to voice also correlated with the BPVS (all $p<0.05$). In the tvFTLD group voice familiarity significantly and strongly correlated with tests of verbal and non-verbal IQ ($p<0.05$).

In the AD group, several tests of voice recognition performance significantly positively correlated with tests of episodic memory: voice familiarity correlated with recognition memory for faces and voice identification correlated with recognition memory for words ($p<0.05$). In this group voice identification and naming tasks also significantly positively correlated with landmark recognition ($p<0.05$).
In both groups, voice naming scores showed a significant relationship with naming of common objects on the GNT ($p<0.05$), and in the tvFTLD group only, voice naming significantly correlated with forwards digit span and reading IQ ($p<0.05$).

Correlations between speaker discrimination tests and background neuropsychological and semantic test scores in each patient group are displayed in Appendix A.4.5. On tests of speaker discrimination, in the AD group easy speaker discrimination significantly positively correlated with forwards digit span, whereas in the tvFTLD group difficult speaker discrimination scores significantly correlated with performance IQ. No other significant correlations with voice perceptual scores were found.

4.3.5. Individual patient data

In order to assess variation in performance on perceptual and semantic measures individual subject data was examined for each patient group. Performance on each test was classed as impaired if below the 5th percentile cut-off score for the healthy control group.

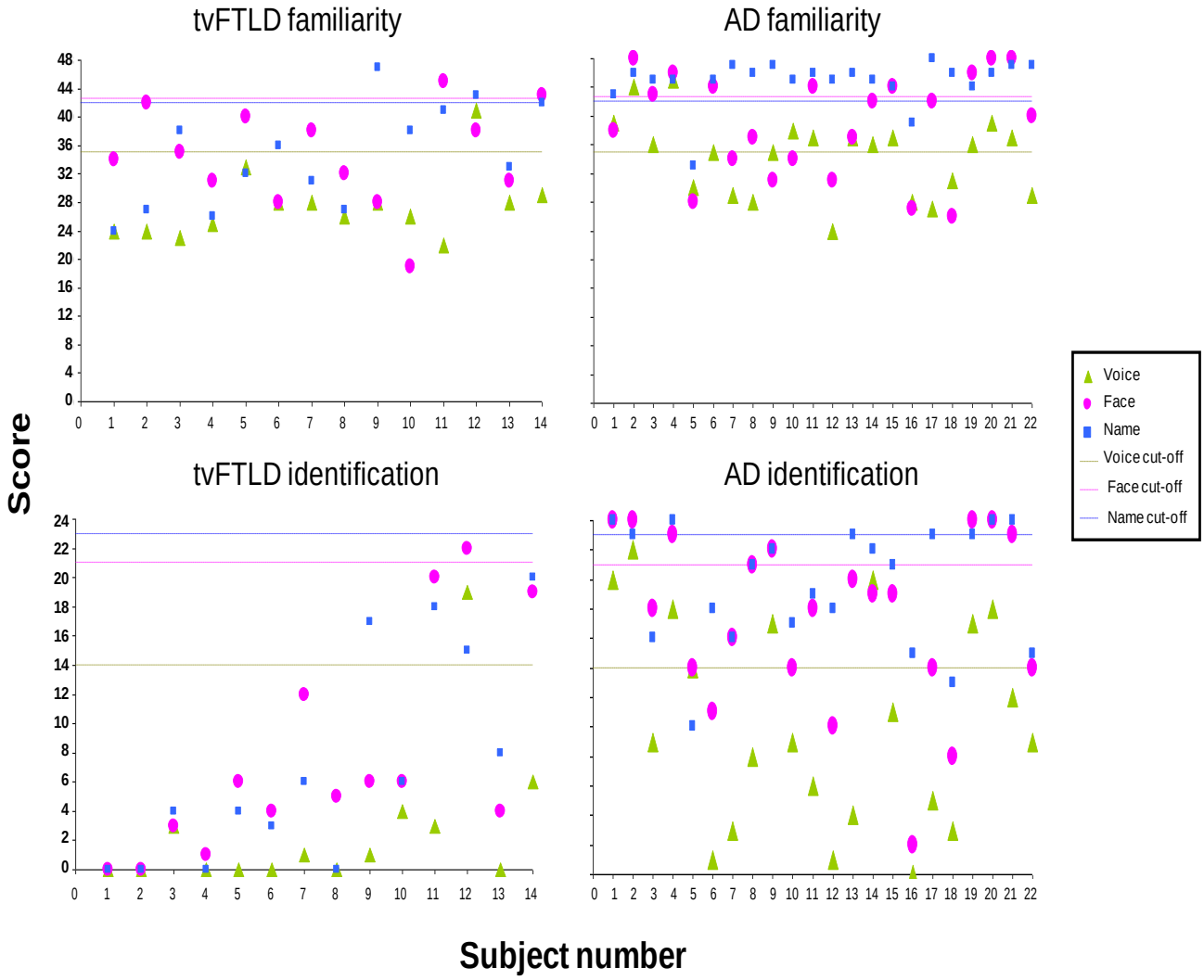
Perceptual tests

On tests of voice perception, 7/22 AD patients but no tvFTLD patients showed impaired performance on one task or more. Four AD patients fell into the impaired range on the gender discrimination test, the remaining patients scoring at ceiling on this task. On the vocal size subtest none of the patients fell into the impaired range due to the large range of control scores. On the speaker discrimination tasks, five patients in the AD group but no patients in the tvFTLD group performed below the 5th percentile of healthy control scores. On the Benton face perceptual matching task, in both patient groups, a similar proportion of individual subjects (5/14 tvFTLD; 9/22 AD) were impaired according to standardised criteria; however, 6/9 AD patients but no tvFTLD patients met criteria for severe impairment. Four AD subjects were impaired at one or more voice tasks but not at perception of faces on the Benton, and three subjects were impaired in both modalities.

Recognition tests

Individual subject data for each patient group showing familiarity and identification performance in each modality (in relation to 5th percentile cut-off scores from the healthy control group) are presented in Figure 4.2.

Figure 4.2. Individual patient data for voice, face and name semantic subtests



Patients with tvFTLD are ordered by performance on a general semantic measure (BPVS score); patients with AD are ordered by a measure of clinical severity (MMSE score). Green triangles show individual data for voice subtests; pink circles show data for face subtests; blue squares show data for name subtests. Dashed lines show 5th percentile cut-offs for each modality calculated from control data.

Appendix A.4.6 shows the number of patients in each group classed as impaired on each modality of presentation for each semantic task (familiarity, naming, identification and cross-modal

matching). Deficits of voice familiarity generally co-occurred with deficits in at least one other modality in both patient groups, most commonly with faces. No patient in either group showed selective sparing of voice familiarity; one patient in the tvFTLD group showed an isolated deficit of voice familiarity, however his peripheral hearing was not normal (see Section 4.2.1). In the tvFTLD group, most (8/14) patients performed best for face familiarity judgments with four patients showing significantly better performance for voice and face familiarity than name familiarity (Revised test of difference $p<0.05$). The AD group showed the reverse pattern with most (16/22) patients performing best for name familiarity judgments, with eight patients showing significantly better performance for name familiarity than voice and face familiarity (Revised test of difference $p<0.05$). Voice deficits in this group in all cases co-occurred with impairments in the face modality, however impairments of face processing were more common: five patients showing an isolated deficit of face familiarity and nine patients performed significantly better for voice familiarity than face familiarity (Revised test of difference $p<0.05$).

Naming, identification and cross-modal matching deficits affecting all modalities were seen in all but one tvFTLD patient, and were present in approximately half of AD patients, the remainder of this group showed substantial heterogeneity of performance across modalities.
In both disease groups, a high proportion of individual cases impaired on tests of voice recognition (13/14 tvFTLD, 11/22 AD) had no perceptual deficit; 4/22 AD patients (but no tvFTLD patients) showed a perceptual deficit in addition to a semantic deficit.

4.3.6. Right versus left temporal lobe damage in tvFTLD

Within the tvFTLD group, nine patients (all with a clinical syndrome of SD) had predominantly left temporal lobe atrophy, while four patients (three with SD, one with bvFTD) had predominantly right temporal lobe atrophy. Demographic characteristics and neuropsychological and experimental test performance profiles of the subgroups across modalities on the semantic subtests characteristics of the two subgroups are compared in Table A.4.7 in Appendices. The two subgroups did not differ significantly in age, years of education or disease duration. The right temporal subgroup had significantly higher MMSE and verbal IQ scores than the left temporal subgroup; the two subgroups did not differ significantly on other general neuropsychological measures. On the experimental measures, the right temporal lobe subgroup performed better than the left temporal lobe subgroup on tests of perceptual analysis of voices, on the Benton task, and on naming tests in all modalities; whereas the left temporal lobe subgroup performed better than the right temporal lobe subgroup on tests of voice and face familiarity and identification. The

difference in performance between the two subgroups was significant ($p<0.05$) only on the Benton task. Consistent with previous structural and functional imaging work (Demonet *et al.*1992; Vandenberghe, Price, Wise *et al.* 1996) these data collectively suggest that the left temporal lobe is integral to verbal semantic memory processes while the right temporal lobe plays a greater role in non-verbal semantic memory processes. Semantic processing of voices may preferentially segregate with semantic processing of faces rather than names. However, the small subgroup sizes indicate a need for caution in extrapolating these data.

4.4. Neuroanatomical data

4.4.1. Neuroanatomical correlates of experimental tests

Interactions between disease group and performance

No significant grey matter associations were identified for group-performance interactions for any of the experimental tests 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 MNI coordinates: 35 -51 -6; cluster size 123 voxels, $p<0.05$ after FWE correction). Voice discrimination performance in the AD group (but not the tvFTLD group) was positively associated with grey matter in right inferior parietal cortex ($p<0.05$ after FWE correction over the pre-specified small volume of interest; see Table 4.5); 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; see Figure 4.3).

Associations of performance across disease groups

The results of the neuroanatomical analysis across both the tvFTLD and AD groups (adjusting for group membership) are summarised in Table 4.5; SPMs are presented in Figure 4.3.

Table 4.5. VBM data: neuroanatomical associations of experimental test performance

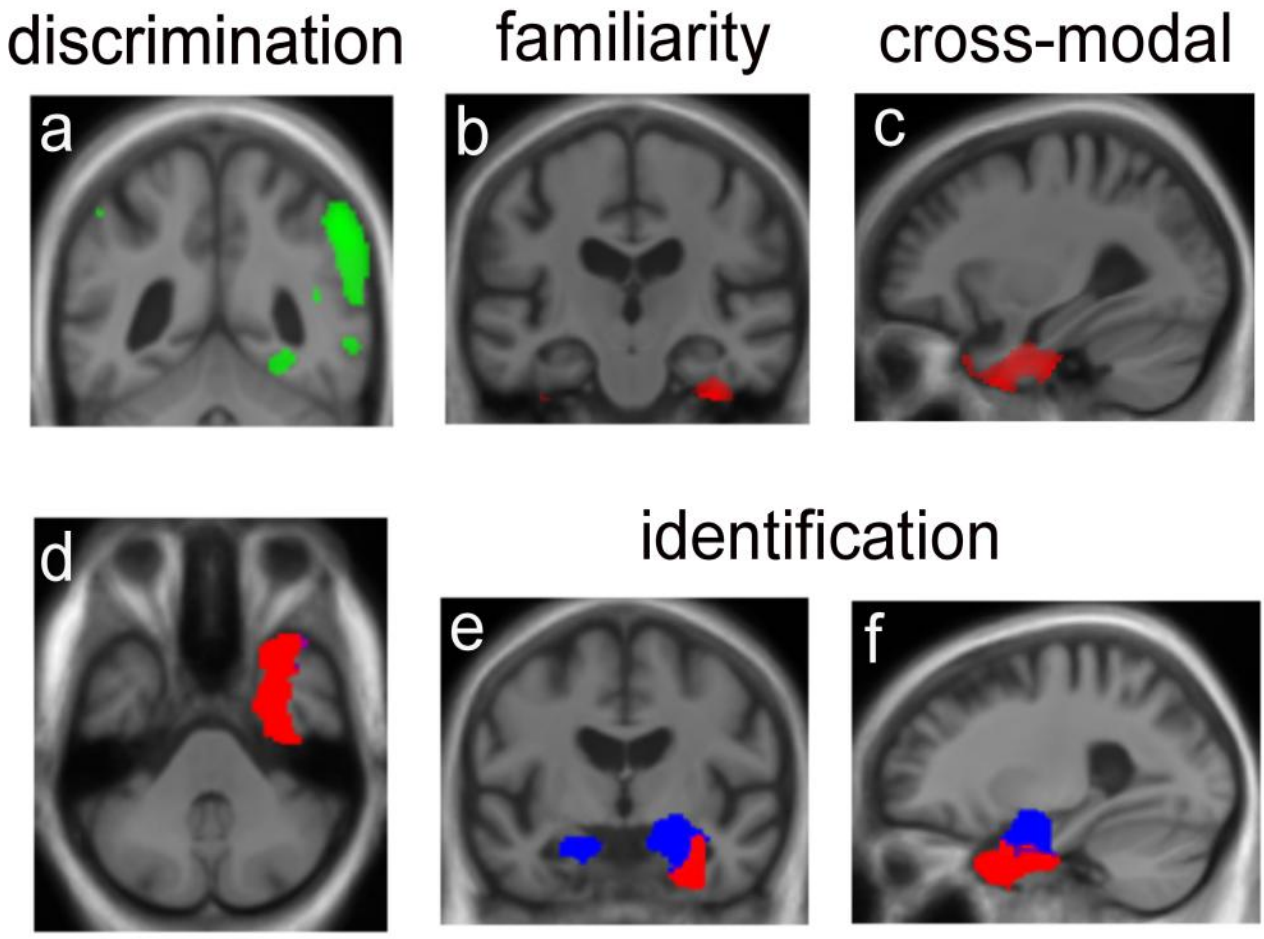
Task	Neuroanatomical associations						
	Side	Area	Cluster size (voxels)	Local maxima			
				Coordinates (mm)			Z score
Apperceptive							
Voice discrimination[a]	R	Inferior parietal cortex	838	56	-50	42	4.13
Familiarity							
Face	L	Temporal pole	678	-55	5	-42	4.29
		Anterior middle temporal gyrus		-62	2	-30	4.11
	R	Anterior fusiform gyrus	250	28	4	-53	4.20
Voice, face & name†	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 & name†	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 & face	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 AD group only; all other results were derived across the tvFTLD and AD groups. All clusters of size >10 voxels are presented. [a] 'easy' version of the speaker discrimination task (see text). †results based on combined-modalities analyses; other results based on separate-modality analyses (see text). *areas with local maxima exceeding a voxel-wise significance threshold $p<0.05$ after FWE correction over the whole brain; other local maxima after correction over the prespecified small volume of interest (coordinates in MNI stereotactic space).

Firstly the results of analyses for associations of experimental test performance over the whole brain volume were considered. No significant associations of voice perceptual performance across both disease groups were identified. In the separate-modality analyses of semantic processing of person knowledge, cross-modal recognition of voices, and identification and cross-modal recognition of faces was 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 was each positively associated with grey matter volume in right anterior fusiform gyrus (all $p<0.05$ after FWE 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 FWE 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 FWE 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 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.

Figure 4.3. Statistical parametric maps of grey matter volume associated with voice processing performance



SPMs show grey matter associations of experimental test performance across the tvFTLD and AD groups (except **a**, see also Table 4.5): (**a**) speaker discrimination (AD group only), (**b**) voice familiarity, (**c**) cross-modal matching of familiar voices and faces, and (**d** – **f**) voice identification (all for tvFTLD and AD 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). SPMs are presented on sections of the mean normalised T1-weighted structural brain image in DARTEL space. Coronal (**a**,**b**,**e**), axial (**d**) and sagittal (**c**,**f**) sections are shown, targeting the 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 SPMs 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 (see Table 4.5); SPMs are thresholded at $p<0.001$ uncorrected for display purposes.

4.4.2. Neuroanatomical correlates of general semantic tests

Interactions between disease group and performance

No significant grey matter associations were identified for group-performance interactions for any of the experimental over the whole brain volume.

Associations of performance across disease groups

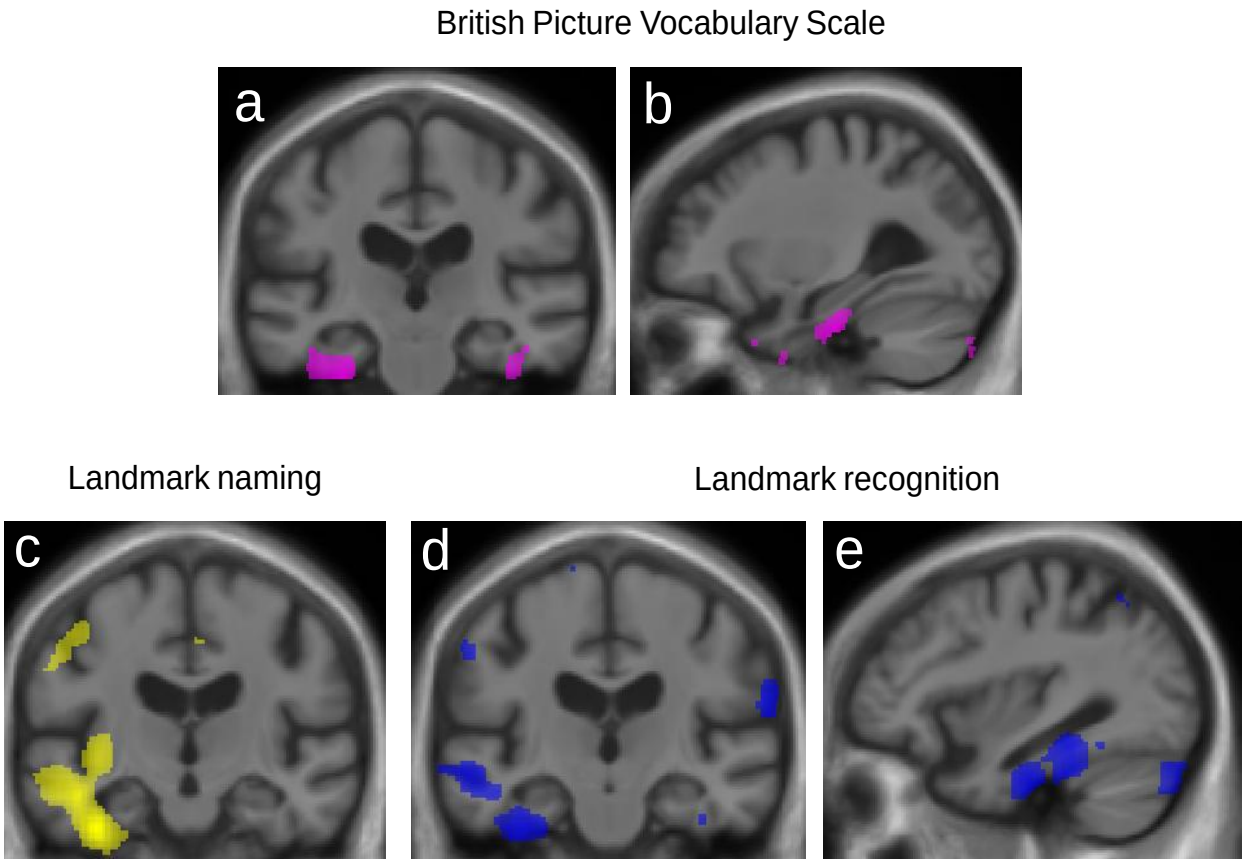
The results of the neuroanatomical analysis across both the tvFTLD and AD groups (adjusting for group membership) are summarised in Table 4.6; SPMs are presented in Figure 4.4.

Table 4.6. VBM data: neuroanatomical associations of general semantic test performance

Task	Neuroanatomical associations						
	Side	Area	Cluster size (voxels)	Local maxima			
				Coordinates (mm)			Z score
Semantic tests							
BPVS	L	Inferior temporal gyrus	3328	-60	-31	-27	4.63
		Parhippocampal gyrus/fusiform		-30	-22	-36	4.20
Landmark naming	L	Anterior fusiform**	99	-39	-15	-35	5.21
		Anterior MTG**	55	-58	5	-25	5.01
		STS/MTG*	127	-58	-36	-7	4.98
		Temporo-occipital junction*	75	-50	-59	-13	4.84
		Posterior fusiform/parahippocampal gyrus*	40	-37	-38	-18	4.83
		STG	15657	-56	-4	-15	4.54
		Insular		-36	-5	-7	4.28
Landmark recognition	L	Anterior MTG*	15	-58	7	-28	4.93
		Posterior fusiform	4440	-41	-37	-31	4.10
		Anterior fusiform		-37	-19	-34	4.06
	R	Posterior fusiform	869	38	-28	-31	4.31
		Anterior fusiform	120	25	-1	-50	4.09

Results were derived across the tvFTLD and AD groups. All clusters of size >10 voxels are presented. *areas with local maxima exceeding a voxel-wise significance threshold $p<0.05$ after FWE correction over the whole brain; **areas exceeding $p<0.01$ FWE correction over the whole brain; other local maxima after correction over the prespecified small volume of interest (coordinates in MNI stereotactic space).

Figure 4.4. Statistical parametric maps of grey matter volume associated with semantic task performance



SPMs show grey matter associations of semantic test performance across the tvFTLD and AD groups (see also Table 4.6): (**a-b**) BPVS, (**c**) Landmark naming (**d-e**) Landmark recognition (all for tvFTLD and AD groups combined). SPMs are presented on sections of the mean normalised T1-weighted structural brain image in DARTEL space. Coronal (**a**,**c**,**d**) and sagittal (**b,e**) sections are shown, targeting the anterior and inferior temporal lobes. The sagittal sections are derived from the left hemisphere and the left hemisphere is shown on the left in all other sections. All SPMs 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 (see Table 4.6); are thresholded at $p<0.001$ uncorrected for display purposes.

Over the whole brain volume landmark naming and identification tests were associated with left anterior MTG, additionally naming was association with anterior fusiform gyrus, and more

posterior temporal regions including STS/MTG, fusiform and temporo-occipital junction in the left hemisphere (all $p<0.05$ after FWE correction over the whole brain volume). Restricting analyses to the pre-specified anatomical volumes of interest, a number of additional associations were identified (all $p<0.05$ after FWE correction over the relevant small volume). Across both disease groups BPVS was positively associated with grey matter in the left inferior temporal, fusiform and parahippocampal gyrus; landmark naming was additionally positively associated with grey matter in the left STS and insula, while landmark recognition was positively associated with grey matter bilaterally in posterior and anterior fusiform gyri. No significant grey matter associations with concrete and abstract synonyms tests were identified.

4.5. Discussion

4.5.1. Neuropsychological impairments of voice processing in tvFTLD and AD

Relative to healthy control subjects, both disease groups were impaired on several measures of semantic processing of voices (voice familiarity, identification, naming and cross-modal identity matching) and voice naming was particularly severely affected; the tvFTLD group performed significantly worse than the AD group across semantic subtests. A qualitatively similar pattern of deficits to voices was observed for semantic processing of faces, while the largest discrepancy between the disease groups occurred for the processing of personal names, particularly at the familiarity task. 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 (particularly recognition of faces). Individual performance profiles corroborated the findings from the group analysis: a substantial proportion of patients in both disease groups showed deficits of identification in all three modalities and a substantial proportion of the tvFTLD group also showed deficits of familiarity in all three modalities. On measures of perceptual processing of voices, the AD group (but not the tvFTLD group) showed deficits relative to healthy controls, and voice apperceptive performance correlated with voice familiarity but not with face apperceptive performance. Taken together, these behavioural data suggest that impairments of voice recognition are significant in both these canonical dementia syndromes but particularly severe in tvFTLD, whereas impairments of voice perception may show relative specificity for AD.

Impairments of voice recognition have rarely been reported in FTLD or AD and may not be clinically salient symptoms due to a lack of assessment of auditory impairments diagnostically

and due to the availability of compensatory cues (from other modalities and contextual cues). However a substantial proportion of the carers of patients in both groups (50% AD, 64% tvFTLD) noticed that the patient had some difficulty with voice recognition. The behavioural findings here provide further evidence that familiarity judgements and cross-modal matching (which do not require explicit access to a person's name or biographical details) are nevertheless sensitive to semantic memory breakdown (Gainotti 2007b; Hanley *et al.*1998; Snowden *et al.*2004). Person recognition models predict that familiarity decisions occur at a post-perceptual level of the PINs: the modality-free gateway to the semantic system (Belin *et al.*2004; Burton *et al.*1993; Burton *et al.*1990). However, analogously with familiarity for other kinds of stimuli, it is likely to represent a multi-component cognitive operation with perceptual, affective, executive as well as semantic dimensions (Gainotti 2007b; Lucchelli *et al.*2008). In the tvFTLD group voice familiarity correlated with performance and verbal IQ and potentially reflects a role for frontal executive components in familiarity judgements as proposed by others (Gainotti 2007b; Lucchelli *et al.*2008). This multi-dimensionality might account for the more variable deficits of familiarity observed within and between disease groups here. Whereas there was strong evidence that voice semantic sub-test performance (for naming, identification and cross-modal matching tasks) correlated between modalities (in particular faces), this was not true for familiarity (notably in the tvFTLD group). 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 VBM (Olson *et al.*2007).

4.5.2. Neuroanatomical correlates of familiar voice recognition impairments in tvFTLD and AD

The neuroanatomical analysis provides support for a common brain mechanism in the right ATL that is critical for voice recognition as well as other modalities of person knowledge: recognition performance across modalities, subtests and disease groups was associated 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 (Andics *et al.*2010; Belin *et al.*2003; Belin *et al.*2002; Belin *et al.*2000; Nakamura *et al.*2001; Shah *et al.*2001; von Kriegstein *et al.*2003; von Kriegstein *et al.*2005; von Kriegstein *et al.*2004) and are likely to be critical for other aspects of semantic processing across sensory modalities (Acres, Taylor, Moss *et al.* 2009; Drane, Ojemann, Aylward *et al.* 2008; Mion *et al.*2010; Williams, Nestor, & Hodges 2005). Indeed, it has been proposed that the temporal pole acts as a pan-modal (or amodal) hub in the semantic processing hierarchy (Lambon Ralph *et al.*2008). These same areas are sites of heavy

disease involvement in tvFTLD, consistent with the more severe person recognition deficits in this group compared with the AD group. However, it is unlikely the profile of anatomical associations observed was driven entirely by the tvFTLD group since there was no evidence of a significant difference in grey matter associations of semantic test performance between the disease groups. Whereas AD is pathologically homogeneous, 'tvFTLD' is likely to be pathologically heterogeneous (Josephs, Hodges, Snowden *et al.* 2011) therefore a shared macro-anatomical substrate is likely to be underpinned by distinct patterns of cellular involvement and correspondingly distinct pathophysiological mechanisms both between AD and tvFTLD groups and within the tvFTLD group.

Certain mesial temporal lobe structures (amygdala, hippocampus and entorhinal cortex) showed an association here with voice identification but not other modalities of person knowledge at the prescribed threshold (see Table 4.5); the mesial temporal lobe has been previously implicated in familiar voice processing by healthy subjects (Andics *et al.*2010; Nakamura *et al.*2001; von Kriegstein *et al.*2005; von Kriegstein *et al.*2004) and this region may be involved in processing sensory object information (Lee, Buckley, Pegman *et al.* 2005) and particularly in tracking information in sound for example, familiar musical melodies: (Samson & Zatorre 1992; Watanabe, Yagishita, & Kikyo 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 ATL in pan-modal processing of semantic knowledge and more particularly with person knowledge (Bozeat *et al.*2000; Coccia *et al.*2004; Lambon Ralph *et al.*2008; Luzzi, Snowden, Neary *et al.* 2007; Rami, Loy, Hailstone *et al.* 2007), previous functional imaging and lesion evidence suggests that these regions participate in a cooperative network mediating different aspects of semantic processing (Ellis *et al.*1989; Grabowski, Damasio, Tranel *et al.* 2001; Mion *et al.*2010; Papagno & Capitani 1998; Thompson *et al.*2004; Tranel 2006; Tranel, Damasio, & Damasio 1997; von Kriegstein *et al.*2005). Associations were observed across all three modalities of person recognition task with the right anterior fusiform gyrus: a region anatomically located at the end of the ventral visual stream is of interest to models of voice processing. Associations with these "basal temporal areas" (anterior fusiform and inferior temporal gyri) were found also with two other semantic tasks: landmark recognition tasks and the BPVS, tasks that were also impaired in both disease groups here (see Table 4.2). Previous

work has associated basal temporal areas with semantic processing of other objects (visual and verbal) in patients with FTLD and focal temporal lobe lesions (Acres *et al.*2009; Mion *et al.*2010; Williams *et al.*2005)) and also in functional imaging of healthy controls performing semantic tasks (Binney, Embleton, Jefferies *et al.* 2010; Bright, Moss, & Tyler 2004; Martin & Chao 2001; Sharp, Scott, & Wise 2004) suggesting that they are integrally recruited in semantic processing mechanisms which are not specific to the modality of input.

Recent models of voice recognition propose that fusiform regions are recruited via cross-modal connections with auditory perceptual regions in STS/STG (von Kriegstein *et al.*2006; von Kriegstein *et al.*2005). Alternatively a more general role in semantic processing has been hypothesised for this region: it has been proposed that fusiform gyri are the "hub" for "basic level" amodal semantic concepts whereas more specific conceptual knowledge about "unique entities", such as naming people, buildings and towns are represented at the temporal poles (Mion *et al.*2010). Although no direct comparison was made here with tasks assessing basic level semantic concepts, associations with the anterior fusiform gyrus across person recognition tasks and in particular with familiarity judgements across modalities may reflect the operation of person identity nodes in this region. The precise role of the anterior fusiform and temporal pole have not been resolved in this study, the fusiform is likely to interact with other cortical areas involved in perceptual (including cross-modal) analysis, whereas the temporal pole may have a more specific role in activating conceptual knowledge about "unique entities" (such as people or buildings). It is often assumed that voice recognition is normal in prosopagnosia, however a single case study indicates that this may not be the case (von Kriegstein *et al.*2006); further studies of voice recognition performance in acquired prosopagnosia with selective damage to inferior temporal regions are necessary.

4.5.3. Voice recognition and models of semantic memory

Evidence concerning the behavioural and neuroanatomical differentiation of modalities of person knowledge is relevant to the more fundamental question of the organisation of brain knowledge systems. According to the two leading theoretical positions, different modalities of knowledge could feed into a unitary amodal system centred on ATLs or alternatively, modalities might continue to be represented within a multiply interconnected semantic network distributed between the left and right temporal lobes. These theories make divergent predictions about deficits of person knowledge following brain damage. The amodal position would predict that semantic deficits should affect all modalities of knowledge in a correlated manner, and modality-specific

deficits would arise from pre-semantic stages of processing; whereas the multimodal position would predict the existence of semantic-level modality-specific deficits despite intact pre-semantic processes, if damage involves modality-specific components of the distributed semantic network. The results of this study implicate bilateral temporal associations for different kinds of nonverbal semantic processing, including face and landmark recognition.

Rather than a purely amodal or fully multi-modal organisation, the present neuropsychological and neuroanatomical findings suggest that verbal and nonverbal modalities of person knowledge are partially differentiated, whereas different modalities of nonverbal (voice and face) knowledge are more closely aligned. Few individuals exhibited modality-selective deficits (in the AD group a few cases displayed selective deficits of face recognition), however more frequent at the group and individual level was discrepancy between verbal (name) recognition and nonverbal recognition of voices and faces. A subgroup analysis comparing neuropsychological performance profiles of FTLD patients with predominantly left versus right temporal lobe atrophy showed a trend in favour of superior naming for cases with predominantly right-sided atrophy and superior voice and face recognition for cases with predominantly left sided atrophy, though differences did not attain significance. It is likely that both the behavioural and the neuroanatomical analyses here were under-powered to detect laterality and modality-specific effects, particularly as temporal lobe atrophy (though asymmetric) was bilateral in all cases, it is possible for example some left tvFTLD cases may have had greater absolute volume loss on the right side e.g. (Brambati *et al.*2009; Mion *et al.*2010).

Differentiation between the hemispheres with respect to the processing of names versus voices (and faces) would be consistent with previous work. Patients with tvFTLD and predominantly left temporal lobe atrophy typically show superior performance for visual over verbal material (e.g.(Lauro-Grotto, Piccini, & Shallice 1997; McCarthy & Warrington 1988; Snowden *et al.*2004)), while patients with predominantly right-sided temporal lobe atrophy are more likely to exhibit clinically significant agnosias for nonverbal material (Edwards-Lee *et al.*1997; Gorno-Tempini *et al.*2004; Miller *et al.*1993; Perry *et al.*2001). Together with evidence from functional imaging studies (Belin *et al.*2003; Demonet *et al.*1992; Scott *et al.*2006), the data collectively suggest that the left temporal lobe is integral to verbal semantic memory processes while the right temporal lobe plays a relatively greater role in non-verbal semantic memory processes. The relations between voice and face processing remain poorly defined and this important issue will only be settled in future studies ideally with larger, anatomically defined clinical cohorts.

4.5.4. Associations with other neuropsychological tests and disease severity measures in tvFTLD

In addition to severe deficits of person knowledge, the tvFTLD group showed severe impairments on other semantic tasks (see Table 4.2) in keeping with pervasive semantic deficits described in SD. However, voice recognition task performance did not correlate consistently with performance on other general semantic tasks; only famous voice naming correlated with more than one semantic test. This result contrasts with studies which show performance between semantic tests in SD is highly correlated across categories and modalities of knowledge (Lambon Ralph *et al.*1999; Lambon Ralph *et al.*2003). Variation in person recognition task performance was not accounted for by differences in disease severity or background neuropsychological measures. It is possible that there may be partial segregation of representations of person knowledge and other categories or levels of semantic knowledge; dissociations between person knowledge and other categories of semantic knowledge has been shown previously in a few lesion and degenerative cases (Ellis *et al.*1989; Evans *et al.*1995; Hanley *et al.*1989; Thompson *et al.*2004). Comparison between recognition of famous voices and nonverbal objects that are of comparable familiarity and frequency within the auditory modality is needed to investigate this further.

Lack of significant correlation between semantic tasks in the tvFTLD group could in principle be due to floor effects on semantic tasks. However voice naming, a task where a number of tvFTLD subjects scored at floor, significantly positively correlated with a number of other tasks: including background semantic tests, naming on the GNT and also forwards digit span and reading IQ. No significant correlation was observed between these tests and other voice semantic tasks. Voice naming performance did not show any significant relationship with any other voice recognition task, and deficits of voice naming may have a different neuroanatomical basis to other voice semantic tasks in this group; for example impairments of naming of both people and common objects have been associated with atrophy in the left ATL in SD (Giovanello, Alexander, & Verfaellie 2003; Mesulam, Rogalski, Wieneke *et al.* 2009).

4.5.5. Neuropsychology and neuroanatomy of vocal apperceptive deficits

Whereas the profile of voice recognition impairment was consistent across subtests and disease groups, deficits of voice perception were restricted to the AD 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 was identified in the right inferior parietal lobe, and there was some

evidence of correlation between voice apperception and recognition performance in the AD group. Taken together, 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 AD group though the relations between perceptual and semantic deficits were not uniform for individual AD patients.

Cognitive models postulate that pre-semantic structural encoding processes derive 'view invariant' representations of both voices and faces which are subsequently linked with stored representations of individual identity in face or voice recognition units (Belin *et al.*2004; Bruce *et al.*1986; Burton *et al.*1990). The present neuroanatomical data do not resolve the mechanism of the perceptual deficit: functional imaging work in healthy subjects has emphasised the role of posterior regions of the STS in encoding complex sound attributes relevant to voice analysis (Belin *et al.*2000; Menon *et al.*2002; von Kriegstein *et al.*2006; Warren *et al.*2005), while neuropsychological evidence has suggested that the right parietal cortex is critical for voice processing (Van Lancker *et al.*1988; Van Lancker *et al.*1989; von Kriegstein *et al.*2006). Inferior parietal lobe activations have been found in healthy listeners as part of a network of regions implicated in voice processing under non-canonical listening conditions (Bishop *et al.*2009). Inferior parietal cortex may therefore be involved in the structural representation of voices, perhaps by holding voice information on-line in working memory for comparison with incoming alternative views of the speaker (e.g., the same voice speaking different phonemes). Weak correlations between speaker discrimination performance and a measure of auditory verbal working memory: forwards digit span, suggests that voice apperceptive deficits may co-occur with or be partially determined by auditory verbal or phonological working memory impairments in AD. Together these suggest testable neuropsychological and neuroanatomical hypotheses for future work.

4.5.6. Associations with other neuropsychological tests and disease severity measures in AD

Perceptual impairments are unlikely to account entirely for deficits of voice recognition in AD: the existence of semantic deficits in other modalities is well documented (Lambon Ralph *et al.*2003; Perry *et al.*2000) and the AD group here showed evidence of more general semantic impairment on background neuropsychological testing (albeit less severe than in the tvFTLD group), and correlations between voice recognition performance and other semantic tasks (face

identification and landmark recognition). In this series, voice recognition impairments never occurred in isolation, for example voice familiarity impairments always co-occurred with impairments in the face modality, whereas isolated deficits of face recognition were observed in five patients. This raises the possibility that vocal processing deficits develop after the onset of facial recognition deficits in AD. This hypothesis could be tested in patients in the early stages of AD in a future longitudinal study.

In the AD group there was substantial variation in performance on the experimental tasks: several patients fell into the unimpaired range on both familiarity and identification tasks. There was not a clear relationship in this group with disease severity measures or semantic tasks. Other potential contributions to voice recognition performance in AD need consideration: it has been proposed that episodic memory about individuals contribute importantly to face recognition (McCarthy & Warrington 1992) and the present data show significant correlation between voice recognition (familiarity and identification of voices) and episodic memory tasks (both verbal and visual recognition memory measures), a result that was not found in the tvFTLD group. This is unlikely to be a coincidental correlation between voice recognition and episodic memory as a result of disease severity. The result might be at least partly attributable to the identification protocol here which encouraged subjects to provide biographical events associated with famous individuals, though this interpretation would be more difficult to sustain for familiarity judgments. This result may be a neuropsychological indicator of medial temporal involvement in voice recognition tasks as suggested by the neuroanatomical data. Impairments of semantic memory for familiar people in AD may cluster with episodic recall for words and faces (Stopford *et al.*2008).

4.5.7. Voice performance and models of voice processing

Unlike functional imaging studies of voice recognition (Andics *et al.*2010; Belin *et al.*2003; Belin *et al.*2000; Nakamura *et al.*2001; von Kriegstein *et al.*2004) associations were not found with superior temporal regions in anterior or mid STS/STG, regions that are thought to be involved bilaterally in voice-specific representations. This may reflect the rarity of voice-specific deficits in either group at semantic and perceptual tasks; across both groups only one subject with tvFTLD performed significantly worse at a voice semantic task (familiarity) when compared to performance in the other modalities, which may have been due to peripheral hearing difficulties, only 4 out of 22 AD subjects were impaired on vocal perceptual tests but not at the Benton. In a future study isolation of the auditory or neuropsychological deficits underlying selective vocal perceptual deficits requires investigation.

In this study associations with voice identification were found in the right amygdala; activations here have been shown in familiar voice recognition tasks (Andics *et al.*2010; von Kriegstein *et al.*2004), for example von Kriegstein and Giraud found that the right amygdala showed functional connectivity with anterior superior temporal regions during recognition of familiar voices (von Kriegstein *et al.*2004). Cognitive models of voice processing (Belin *et al.*2004) propose that there is parallel processing of identity information and emotional content from voices, on the basis of evidence from neuroimaging and a neuropsychological study of developmental phonagnosia (Garrido *et al.*2009; Imaizumi *et al.*1997) however it is possible that these two pathways in the voice modality are not fully independent, consistent with psychophysical and electrophysiological evidence (Campanella & Belin 2007). It has been shown that musical instrument timbres carry an affective tone (Hailstone *et al.*2009; Juslin & Laukka 2003), for example the violin timbre biased healthy subjects towards "sadness" judgements on melodies in one study (Hailstone *et al.*2009), and it may be that the affective quality of an individual's voice contributes identity information.

4.5.8. Methodological considerations

This study has several methodological considerations. A number of patients (primarily in the tvFTLD group) performed at or near floor on voice tasks, and it is possible that voice recognition was too difficult. Across patient and control groups, voice recognition was the most difficult and name recognition tasks was the easiest, a pattern of results expected on the basis of previous work. As the identities of the public figures used were the same in all three modalities this may have amplified this discrepancy, as no attempt was made to match control recognition frequency across modalities and also as the same individuals were tested, priming is likely to have occurred in the face and name modalities (presented after voices) which may have artificially improved performance, particularly in controls. Impairments of voice processing at the level of individual patients or within each patient group however were not disproportionate to deficits in other modalities relative to controls. Perhaps more problematic than low performance on voice tests in patients is ceiling control scores for several face and name semantic tasks, which meant that it was difficult to assess patterns of performance across modalities relative to controls. For example it is unclear whether name familiarity scores were relatively preserved in the AD group relative to controls (as has been found previously (Greene *et al.*1996)), whether tvFTLD performance was relatively worse (which has also has been shown previously (Snowden *et al.*2004)) or whether both factors may have been operating. A future study assessing recognition of famous people across modalities using a set of famous people chosen so that each modality is matched for

control recognition frequency could be achieved by using less common celebrities in face and name modalities (as in Study 1 in this thesis), or by making faces less recognisable, for example by blurring them (Damjanovic *et al.*2007), though each of these paradigms is likely to entail other methodological considerations, particularly when assessing patients with other cognitive or perceptual difficulties. It may be however that structure of person recognition mechanisms is such that it is highly dependent on the face channel (von Kriegstein *et al.*2006; von Kriegstein *et al.*2005), and voices are intrinsically more difficult to recognise than faces (Damjanovic *et al.*2007; Hanley *et al.*1998). This potentially needs to be assessed using the voices and faces of people that are personally familiar to subjects so that recognition is not influenced by channels of exposure in the media.

Another limitation of this study was that despite using an unbiased automated whole-brain study of brain morphometry to investigate the neural correlates of voice processing, only voice cross modal matching VBM results survived whole-brain correction: grey matter associations of voice identification and perception were significant only in ROI analyses. Findings were therefore biased by focussing on the temporal lobes, and associations in other cortical regions implicated in studies of voice processing such as inferior frontal, parietal or cingulate cortices may not have been found due to exclusion in analyses. However, in whole-brain analyses not presented in this thesis, significant associations with all voice familiarity, identification and cross modal tasks were found in the right ATL after FWE correction when group membership was not included as a covariate further justifying ROI analysis in the temporal lobes.

The VBM results indicate the regions of atrophy that correlate with a decrease in score in these patient groups, but they do not directly show which regions are critical to voice processing. Consistent associations across semantic tasks with anterior and interior temporal regions for example may represent a coincidental correlation in which more advanced disease results in greater atrophy in these regions and poorer task performance, which is plausible in the tvFTLD group given that atrophy is concentrated in these regions. However there was limited evidence for the influence of disease severity on experimental test scores, and as described above, group membership was included as a covariate in the analyses which adjusted for the overall mean difference in test performance between the two patient groups. Also identical patterns of association were not found for all semantic tests: voice identification was associated with right-lateralized anterior and medial temporal but not inferior temporal or fusiform regions whereas landmark naming was associated with a large number of left lateralized regions (including

temporo-occipital and insular regions). Furthermore there was no evidence that the semantic test results was driven entirely by the tvFTLD group, as no significant disease specific associations of semantic tasks were found. This may have been because there were insufficient tvFTLD subjects in the VBM analyses (only 11 subjects were included) to detect tvFTLD-specific effects. Alternatively the result may, as presented earlier in the discussion (Section 4.5.2), reflect overlap in the neuroanatomical basis of impairments in the two patient groups given that semantic deficits (including person recognition tasks) also occurred in the AD group, and damage to the anterolateral temporal lobe has been previously associated with semantic impairments in AD (Balthazar, Yasuda, Pereira *et al.* 2010). Investigating voice processing in a study with a larger tvFTLD sample may elucidate this.

4.5.9. Conclusions & future work

In this study the behavioural and neuroanatomical signatures of voice processing deficits in FTLD and AD are described. Deficits on all aspects of voice recognition and impairment were found in both disease groups, but were more severe in the tvFTLD group than the AD group. The AD group showed additional deficits of vocal gender perception and voice discrimination. The VBM analysis across both disease groups revealed that the right ATL is likely to have a critical role in recognition of voices and other modalities of person knowledge. Common grey matter associations of familiarity, identification and cross-modal recognition in all modalities in the right temporal pole and anterior fusiform gyrus were found in combined analyses across the patient groups; while in the AD 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 tvFTLD, whereas impairments of voice perception may show relative specificity for AD. Further analysis of the auditory perceptual impairments underlying deficits of vocal perception in AD at the group or individual level could be undertaken in a future study. 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 customised stimuli that allow particular vocal attributes to be isolated and manipulated.

In addition to deficits of familiar voice recognition, face and name recognition were also impaired in both groups and name recognition was significantly more impaired than other modalities in the tvFTLD group. Although the statistical analysis adjusted for differences in control performance across modalities, control performance on name (and face) semantic tasks were at or close to

ceiling, and a future study matching control recognition frequency across modalities (rather than matching identities of public figures across modalities) may be useful for establishing the pattern of performance across modalities in the two patient groups. There was some evidence from neuropsychological and neuroanatomical data that nonverbal objects (faces and voices) and verbal entities (names) show partial segregation in the right and left temporal lobes. A larger cohort, in particular of tvFTLD subjects will be necessary to determine any laterality or modality-specific effects neuropsychologically, for example by comparison of a larger sample of right and left sided tvFTLD cases, while neuroanatomically, VBM analyses may have been underpowered to detect any modality specific associations.

The tvFTLD group showed severe deficits both at person recognition tests at other semantic tasks: landmark recognition (recognition of visual unique entities) and a test of general semantic knowledge (BPVS): neuroanatomical associations with these tasks in the anterior and inferior temporal lobes were similar to person recognition tasks. Performance on these tests however did not correlate with voice recognition performance, suggesting partial segregation between modalities, categories or levels of knowledge. A neuroimaging study directly comparing semantic tasks assessing recognition of less differentiated versus highly differentiated auditory objects (such as recognition of environmental sounds versus musical melodies, which may be a highly differentiated auditory category in healthy people with some musical experience), may establish whether there is differential organization of these concepts for example in inferior versus polar regions of the temporal lobe as hypothesized by others (Mion *et al.*2010). Associations with the anterior fusiform gyrus across modalities of presentation for familiarity (and other person recognition tests) implicate inferior temporal regions in formation of multimodal representations of familiar people and other object identities. Investigation of voice recognition in acquired cases of prosopagnosia with damage to inferior temporal lobes may further illuminate the role of "basal temporal" regions in processing of familiar voices. A larger sample size of tvFTLD subjects may clarify the extent to which anterior and inferior temporal associations of semantic processing are specific to tvFTLD.

This study suggests clear directions for future work. It has been proposed that modality-specific deficits of person knowledge become generalised with the evolution of neurodegenerative disease (Evans *et al.*1995; Gainotti *et al.*2003; Gainotti *et al.*2008; Gentileschi *et al.*1999): the present study suggests that deficits of person knowledge are not uniformly related to disease duration or severity, and therefore the profile of development of these deficits may hold information about

the organisation 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. It will be important in future studies to directly compare vocal identity to emotional processing behaviourally, and may be particularly relevant to relate to measures of social and behavioural change in FTLD. There is a need for correlation of voice processing measures with structural and functional anatomical data and with tissue histopathology in a wider range of neurodegenerative diseases.

5. Study 3: Neuropsychological and neuroanatomical analysis of accent processing in PNFA and AD

5.1. Introduction

Accent processing in the healthy brain and the effects of brain damage on accent processing have been little researched compared to the mechanisms of voice perception and recognition, investigated in Studies 1 and 2. Studies assessing accent processing generally entail two broadly complementary tasks: processing of the accent as an informative vocal signal in its own right (Adank *et al.*2012; Berman *et al.*2003; Clopper *et al.*2004a; Clopper *et al.*2004b; Van Bezooijen *et al.*1999), and processing the effects of the accent on the prototypical speech signal (Adank *et al.*2009; Best *et al.*2001; Clarke *et al.*2004; Evans *et al.*2004; Floccia *et al.*2009; Floccia *et al.*2006). The effects on accent processing of neurodegenerative disease remain largely unknown.

There are grounds to anticipate deficits of accent processing in the canonical degenerative dementias. Disease in AD and PNFA syndromes affects large-scale brain networks including superior temporal, prefrontal and parietal regions implicated in accent processing (Adank *et al.*2012; Berman *et al.*2003; Gorno-Tempini *et al.*2004; Neary *et al.*1987; Rohrer, Ridgway, Modat *et al.* 2010; Scahill, Schott, Stevens *et al.* 2002; Seeley *et al.*2009; Sonty *et al.*2007; Zhou *et al.*2010). In particular, these diseases target anterior (in PNFA) and posterior (in AD) peri-Sylvian cortices that mediate different levels of processing of complex verbal and nonverbal sounds, including inferior parietal regions implicated in apperceptive levels of analysis of voices in the previous study (Chapter 4). Impaired processing of complex nonverbal auditory patterns (Baird *et al.*2009; Eustache *et al.*1995; Goll *et al.*2010; Goll *et al.*2012; Rapcsak *et al.*1989; Uttner *et al.*2006) and other meta-linguistic components of the speech signal, including prosody (Allender *et al.*1989; Horley, Reid, & Burnham 2010; Rohrer *et al.*2010; Taylor *et al.*1971; Testa *et al.*2001), and speaker identity in the previous study (Chapter 4) have been documented in PNFA and AD.

The contribution of different regions of the temporal cortex are of interest to accent processing; in Study 2 specific regions of the superior temporal lobe were not significantly associated with voice processing whereas there was some suggestion that regions of the mesial temporal cortex may show specificity for voice recognition. It is hypothesised that accent processing may be less dependent on cross-modal mechanisms of analysis than voice identity processing (described in

Section 1.5.3) and hence may not recruit inferior regions of the temporal cortex. In the previous study anterior temporal cortices were significantly associated with voice recognition performance in tvFTLD and AD, regions that have been previously implicated in vocal and facial identity and emotion processing in dementia (Ellis *et al.*1989; Gainotti *et al.*2008; Joubert *et al.*2006; Omar *et al.*2010; Omar *et al.*2010; Rosen *et al.*2006): damage involving these regions in neurodegenerative diseases may lead to a more general deficit in decoding social signals, including accents.

In this study, the cognitive and neuroanatomical bases of accent processing were investigated in two canonical neurodegenerative dementia syndromes: PNFA and AD. A novel neuropsychological battery was designed to assess these in cognitively impaired patients, addressing two aspects of accent processing: the intelligibility of accented speech (accent comprehension) and recognition of non-native regional and foreign accents (accent recognition). Neuroanatomical associations of behavioural performance were assessed using VBM. It was hypothesised that these dementia syndromes would be associated with separable behavioural profiles of impaired accent processing. It was further hypothesized that accent comprehension and accent recognition performance would have overlapping neuroanatomical associations in the postero-lateral and anterior temporal regions previously shown to be critical for other aspects of vocal signal processing.

Materials and methods

5.2. Subject demographic characteristics and clinical details

Six patients with PNFA and twenty patients with AD diagnosed according to consensus clinical criteria (Dubois *et al.*2007; Neary *et al.*1998) were recruited. 35 healthy older control subjects described in Study 2 (Section 4.2.1) also participated; background data for both healthy control and AD groups were included in the previous study (Section 4.2.2). Demographic and clinical details of subjects are summarised in Table 5.1.

Table 5.1. Summary of demographic and clinical characteristics of patient and control groups

	PNFA n=6		**AD n=20**		**Control n=35**	
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
Males: females	0:6	-	8:12	-	13:22	-
Age (years)	66.0 (6.9)	58-76	66.4 (7.6)	49-79	65 (6.0)	54-79
Education (years)	12.3 (3.3)*	10-17	13.1 (3.4)*	9-20	15.2 (3.3)	11-25
Symptom duration (years)	3.5 (1.3)**	2-6	6.0 (2.4)**‡	4-11	n/a	
MMSE score (/30)	20.0 (4.9)**	14-26	21.6 (4.1)**	14-28	29.4 (0.6)[a]	28-30

*significantly lower than controls ($p<0.05$); ** significantly lower than controls ($p<0.01$); †significantly different to the other patient group ($p<0.05$); ‡significantly longer than the other patient group ($p<0.01$); [a] n=23 controls performed MMSE

Patient and control groups were well-matched for age. 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. Males were under-represented in the PNFA group relative to AD and control groups (although these differences were not statistically significant), and controls had a significantly greater average number of years of education compared to both patient groups (the patient groups did not differ significantly on this measure). The PNFA and AD groups did not differ significantly on one measure of disease severity (MMSE score) but the AD group had a significantly longer mean symptom duration than the PNFA group.

All subjects were native British residents with English as their first language. In order to gather information about their past accent exposure, at entry to the study all subjects completed a brief questionnaire detailing the region of Britain where they had grown up, their recent regional residence and any extended periods (> 6 months) spent outside the United Kingdom. This information indicated that the overall accent exposure of the subject groups was likely to have been similar. All subjects were resident in Southern England at the time of participation in the study. One patient with PNFA had spent her early childhood abroad (Malta) and one patient with PNFA, six patients with AD and nine healthy control subjects had spent their childhood in

another region of Britain; nine patients with AD and ten healthy control subjects had lived outside the United Kingdom for an extended period (most during their earlier adult life).

Eighteen patients in the AD group and all patients in the PNFA group had undergone previous brain MRI; these images were reviewed by an experienced neuroradiologist. Fifteen of the AD patients had disproportionate bilateral hippocampal atrophy and the remainder had generalised cerebral atrophy. Two AD patients were unable to have MRI due to a cardiac pacemaker; computed tomography in one of these patients showed generalised cerebral atrophy. Patients with PNFA showed bilateral but asymmetric peri-Sylvian atrophy (more marked on the left in three cases and on the right in the remaining cases). No patient had radiological evidence of significant cerebrovascular disease.

General neuropsychological assessment

All patients and 19 healthy control subjects had a comprehensive general neuropsychological assessment; 16 control subjects performed a reduced set of tests. Groups were compared using linear regression, adjusting for age, gender, and years of education, with p-values from z-tests using bootstrap standard errors (2000 replicates).

Assessment of peripheral hearing

Most subjects had no clinical history of hearing loss. One PNFA subject had bilateral high frequency hearing loss, assessed on clinical audiometry. One AD patient had mild bilateral high frequency hearing loss. One control subject had mild bilateral high frequency hearing loss, previously confirmed on clinical audiometry. All subjects underwent screening pure tone audiometry, methods described in Chapter 2 (Section 2.2.2); four subjects (one AD patient, two controls) reported unilateral right sided hearing loss and in these subjects the left ear was tested. Group differences in mean response time at each frequency were assessed using linear regression adjusted for age and gender, with p-values from z-tests using bootstrap standard errors (2000 replicates).

5.3. Experimental investigations: Tests of accent processing

Accent comprehension was assessed in two tasks: comprehension of questions spoken in a native British (Southern Standard English) versus a foreign accent, and verification of single words spoken in a foreign accent relative to the native British English accent. Southern Standard English was used as the reference accent as this is widespread in the Greater London area and was

therefore likely to be highly familiar to most participants. Accent recognition was assessed in three tasks, directed to progressively more fine-grained 'levels' of accent knowledge: identification of an accent as British English or foreign, identification of regional British accents and identification of regional English accents. For both the accent comprehension and accent recognition limbs of the battery, additional tests were administered to assess other cognitive capacities relevant to performance on the accent tasks: these additional tests comprised Southern Standard English phoneme discrimination (a measure of phonological processing, relevant to performance on the word verification subtest) and country recognition (a measure of general geographical semantic knowledge, potentially relevant to performance on the accent recognition subtests).

5.3.1. Experimental investigations: Tests of accent comprehension

Accent comprehensibility is influenced by lexical context, familiarity (Adank *et al.*2009; Clopper *et al.*2008) and acoustic (phonological-phonotactic) distance from native speech (Best *et al.*2001; Clarke *et al.*2004; Floccia *et al.*2006); in these respects, other regional accents generally fall closer to native accent than do foreign accents. For this subtest, three foreign accents were chosen which differ from Southern Standard English at the phonological-segmental and prosodic level: General American, Australian and South African. It was hypothesised that British subjects would be more familiar with American and Australian accents (via the media) than with the South African accent though we anticipated that subjects would have been fairly familiar with all the foreign accents selected for this study. All recorded speakers had English as their first language (in order to eliminate any perceptual costs associated with irregular or dysfluent speech of non-English speakers (Floccia *et al.*2006)) and all were female aged 21 – 42 years; foreign accent speakers had all lived in the United Kingdom for less than 9 months. To minimise any effects from individual speech idiosyncrasies, more than one speaker was recorded for each accent (four Southern Standard English, two American from the Mid-West, two South African from Johannesburg and two from Eastern Australia).

Question comprehension

In this subtest, subjects heard 40 short spoken questions (each between 4 and 8 words in length) designed to elicit a one word answer (Prof EK Warrington unpublished; see Appendix A.5.1); each sentence was spoken once in a Standard English accent and once in a foreign accent (either American or South African), yielding 80 trials in total (40 trials for the English accent and 40 trials for a 'foreign' accent). Trials were presented in four divided blocks of 20 trials. For each

sentence, the presentation order was randomly assigned first to the English accent or first to a foreign accent in the first set of 40 trials and assigned to the other accent category (using a different randomisation order) in the second set of 40 trials. The task on each trial was to answer the question (either aloud, or in the case of patients with PNFA, as a written response if preferred).

Word verification

In this subtest, subjects heard 24 spoken monosyllabic words derived from PALPA (Psycholinguistic Assessment of Language Processing in Aphasia) Minimal Pair Discrimination tests (Kay, Lesser, & Coltheart 1992) (see Appendix A.5.2.) each recorded with Standard English, American, Australian and South African accents. Each spoken word was presented twice using each accent (yielding 24 trials for each accent and 192 trials in total), once with the target written word and once with a foil. Stimuli were presented using Superlab version 4 (http://www.superlab.com/). On each trial the subject was instructed to read a written word (target or foil, with equal probability) presented on a computer screen; two seconds later subjects heard the spoken word, and the task was to indicate whether this spoken word matched the written word.

Word foils each contained a single phonetic change compared with the corresponding target word (half contained a change in vowel sound, half contained a change in initial or terminal consonant). The set of words contained a range of vowel and consonant changes; no attempt was made to manipulate confusability under particular accents. In order to enhance any effect of accent on error rates (and/or reaction times in controls) target words selected had an orthographic and phonological neighbourhood greater than 10 (Grainger 1990; Luce & Pisoni 1998) and foils had a CELEX word frequency greater than the corresponding target word (Baayen, Piepenbrock, & Gulikers 1995). 'Neighbourhood' here refers to the number of similar words of the same length generated by changing one letter while preserving letter position; increasing neighbourhood size is associated with increasing lexical decision time. Psycholinguistic data used in this study were obtained using N-Watch (http://www.maccs.mq.edu.au/~colin/N-Watch/).

Trials were presented in eight blocks each containing the set of 24 spoken words (six words spoken with each of the accents); the presentation order of a particular word under each accent was randomised. Patients responded by pointing to 'Yes' or 'No' listed for each trial in a response sheet; control subjects responded by pressing 'Yes' or 'No' on a response box, and their

reaction times were recorded. As no predictions were made about how confusable distractor words were for each accent, only trials in which the target written word matched the spoken word were analysed, yielding a score / 24 for each accent.

Phoneme discrimination

As a measure of phonological processing ability, a total word verification score was calculated for all target words and foils spoken with a Southern Standard English accent (total score / 48). Words and foils were derived from the PALPA Minimal Pairs Discrimination tests (Kay *et al.*1992) (see Appendix A.5.2) each 'target' word differed from the corresponding 'foil' word with respect to one phonetic feature. The task on each trial was to indicate whether the spoken word matched the spoken word. Patients responded by pointing to 'Yes' or 'No' listed for each trial in a response sheet; control subjects responded by pressing 'Yes' or 'No' on a response box.

5.3.2. Experimental investigations: Tests of accent recognition

Foreign accents

In this subtest, subjects were assessed for their ability to identify an accent as native British (Southern Standard English) or foreign. The same set of 40 questions and accents (Southern Standard English, American, South African) used in the question comprehension subtest was re-presented. Maps were used to assist in explaining the task. On each trial the subject was asked 'Is this person from England?', and responded 'Yes' or 'No' verbally or by pointing on a response sheet. If the subject scored <12 on the first block of 20 trials the test was discontinued; scores for the first block and for all four blocks were analysed.

Regional British accents

For this subtest, audio samples each comprising 7-15 seconds of speech representing a Southern Standard English, Irish, Scots or Welsh accent were obtained from accent archives available on the World Wide Web (http://web.ku.edu/~idea/; http://www.bbc.co.uk/voices/; http://www.bl.uk/learning/langlit/sounds/index.htm; accents are listed in Appendix A.5.3.). In selecting the clips, an attempt was made to minimise extraneous lexical cues to accent origin. Six different speakers representing each of the four accents were selected, yielding a total of 24 trials. The task on each trial was to identify the speaker's regional origin in a four-alternative forced choice procedure (England, Ireland, Scotland, and Wales); a map of the United Kingdom and Ireland labelling each region was also presented with which to respond non-verbally if preferred.

Regional English accents

This subtest was designed to exploit the wide variation in English regional accents as an index of more fine-grained semantic processing of accents. Audio samples representing speakers from either the north or the south of England were selected from the on-line accent archive (http://web.ku.edu/~idea/; http://www.bbc.co.uk/voices/; http://www.bl.uk/learning/langlit/sounds/index.htm; accents utilized are listed in Appendix A.5.3.), following the same selection criteria as the regional British accents subtest. 24 audio clips each representing a different speaker from either the North or the South of England were chosen (avoiding the Midlands, in order to reduce ambiguity), yielding a total of 24 trials. The task on each trial was to identify the speaker's regional origin in a two-alternative forced choice procedure (North or South England); a map of England labelling each region was also presented with which to respond non-verbally if preferred.

Country knowledge

As a measure of general geographical knowledge, knowledge of 10 countries (four British, four European and two non-European; see Appendix A.5.4.) was assessed in three subtests: naming from verbal description; naming from maps; and (if the subject was unable to name all 10 countries) recognition of the map corresponding to the spoken name of the country (forced-choice from an array of 10 maps).

5.4. Analysis of behavioural data

5.4.1. Group statistical analyses

To quantify differences between groups (control, AD, PNFA) on each experimental test, linear regression models were fitted to the scores, adjusting for age, gender and years of education. P-values for group differences were found using a z-test with bootstrap standard errors (2000 bootstrap replicates). In order to investigate the performance cost associated with listening to words or sentences presented in a foreign accent on accent comprehension tests, a difference score was calculated for each subject based on their performance on the question comprehension (total score for foreign accents minus total score for Southern Standard English questions) and word verification subtests (mean score for foreign accents minus total score for Southern Standard English words). Differences between groups for these scores were again assessed using linear regression, adjusting for age, gender and education.

5.4.2. Further analyses in the control group

In addition, on the word verification test, to investigate differences in score by accent (Standard English, American, Australian, and South African) in the healthy control group, differences in mean score between each accent and English were calculated and 95% Wald-type bootstrap confidence intervals (2000 replicates) were obtained. A linear regression model was used to estimate differences in mean reaction time between accents, adjusting for word duration. P-values and 95% confidence intervals were again found using z-tests and Wald intervals using bootstrap standard errors (2000 bootstrap replicates).

5.4.3. Correlation analyses in the AD group

The relationship between experimental tests of accent processing and general phonological or geographical semantic skills was investigated within the AD group, it was not explored in PNFA group as the ability to detect correlations in a sample of this size (n=6) is extremely limited. Correlation coefficients between accent recognition scores and tests of country recognition (country naming from description, map naming and map recognition) were estimated with 95% bias-corrected and accelerated bootstrap confidence intervals (2000 replicates). Similarly correlation coefficients between accent comprehension difference scores (Foreign minus English) and phoneme discrimination task performance were estimated.

In addition, in the AD group the influence of background neuropsychological performance on accent processing, and the relations between accent processing performance and voice processing (assessed in Study 2) were of interest and investigated using pairwise correlation analyses using the method described above. In Study 2, both voice apperception and voice semantic test performance was impaired in AD therefore the relations between performance on these tests and tests of accent processing were also analysed. The relationship between accent processing and a sub-set of background neuropsychological tests directed towards potentially relevant cognitive domains were also assessed: these comprised tests of nonverbal IQ, forwards and backwards digit span, semantic processing on the BPVS and executive function on the Stroop. As tests of voice recognition in Study 2 significantly correlated with tests of recognition memory in AD, correlation between voice performance and accent processing task performance were assessed; list of tests correlated with accent tests are displayed in Appendix A.5.6.

5.5. VBM analysis

3T MRI images were acquired for 17 AD patients and 4 PNFA patients. Associations with regional grey matter volume were assessed separately for performance on each of the accent comprehension subtests (entering the difference score for each subtest) and on each of the accent recognition subtests (entering the raw score for each subtest). For each experimental subtest, performance was assessed in a combined-groups analysis (n=21) in which grey matter volume was modelled as a function of the experimental test score, group and the score-by-group interaction, with age, and TIV as covariates. Using this model, grey matter associations with group-performance interactions were tested for each experimental subtest, and any significant within patient group associations were identified. Additionally, a size-weighted average of the associations in the two groups (AD * 17/21 + PNFA * 4/21) was tested, which is analogous to the simple association in a model without the score-by-group interaction.

For each test, grey matter associations were assessed over the whole-brain and within the temporal lobe regional volume of interest; a voxel-wise statistical threshold $p < 0.05$ family-wise-error (FWE)-corrected for multiple comparisons was applied in all analyses. Statistical parametric maps were displayed as overlays on a study-specific template structural brain image. The grey matter segment of the final DARTEL template was affine registered to the a priori grey matter tissue probability map in SPM, and DARTEL coordinates were transformed using the estimated affine mapping to standard stereotactic MNI space.

5.6. Results: Background tests

5.6.1. General neuropsychological performance

Relative to healthy controls, both patient groups showed impairment on tests of IQ, verbal recognition memory, semantic memory tests, working memory, arithmetic and executive function (see Table 5.2) after adjusting for age, gender and number of years of education. In addition, the AD group was impaired relative to controls on tests of face recognition memory and object perception. The PNFA group was significantly impaired relative to controls and relative to the AD group on tasks dependent on speech output (including reading, digit span and Stroop inhibition) and verbal semantic knowledge; no other significant differences between the disease groups were identified.

Table 5.2. General neuropsychological assessment in patient and control groups

	PNFA n=6		AD n=20		Control n=35	
Test (max score)						
IQ	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
Verbal IQ	63.7 (15.1)**‡	55-92	98.3 (16.7)**	67-121	120.8 (9.2)	96-142
Performance IQ	80.2 (12.1)**	70-100	87.9 (16.7)**	62-110	116.8 (11.9)	100-141
Reading IQ[a]	78.0 (18.9)**†	56-103	106.4 (16.4)**	67-128	118.9 (7.4)	96-129
Episodic memory						
RMT (words) (/50)	33.7 (11.7)*	19-47	30.7 (7.5)**	19-47	47.3 (1.8)	43-49
RMT (faces) (/50)	36.5 (5.4)	30-44	34.9 (5.8)**	25-45	42.2 (4.7)	35-49
Semantic tests						
BPVS (/150)	127.3 (18.0)**	101-146	140.9 (12.4)*	106-150	148.1 (1.5)	144-150
GNT (/30)	7.7 (9.5)**	0-23	12.1 (8.1)**	0-26	26.0 (2.4)	19-30
Concrete synonyms (/25)	17.0 (2.9)**†[b]	13-20	20.8 (2.7)**	13-25	24.3 (1.3)	19-25
Abstract synonyms (/25)	17.6 (4.0)**†[b]	12-23	20.9 (3.6)**	14-25	24.3 (1.2)	20-25
Working memory						
Digit span fwd (/12)	4.0 (3.5)**†	0-5	7.5 (2.2)	4-11	8.7 (2.0)	4-12
Digit span back (/12)	1.8 (1.7)**‡	1-9	5.2 (2.7)*	0-10	7.4 (2.6)	2-12
Spatial span fwd (/12)	4.8 (1.3)*[b]	4-7	5.7 (2.4)[c]	1-9	6.8 (1.5)[d]	5-9
Spatial span back (/12)	3.8 (2.0)**[b]	2-6	4.0 (2.0)**[c]	0-7	6.7 (1.7)[d]	4-10
Other skills						
Object decision (/20)	16.8 (1.9)[b]	14-19	15.7 (2.9)**	9-19	18.5 (1.2)	16-20
Arithmetic (GDA) (/12)	3.0 (3.2)**	0-8	5.7 (4.6)**	0-14	15.4 (4.8)	6-23
Stroop switching scaled score (/18)	1.2 (0.4)**‡	1-2	3.9 (3.2)**	1-11	11.5 (2.0)	7-14

p values are for group differences after adjusting for age, gender and years of education * significantly worse than controls ($p<0.05$); ** significantly worse than controls ($p<0.01$); † significantly worse than AD group ($p<0.05$); ‡ significantly worse than AD group ($p<0.01$); WMS-R digit span tests: forwards, backwards; GDA, Graded Difficulty Arithmetic; WMS-III Spatial Span tests: forwards, backwards; WASI, Wechsler Abbreviated Scale of Intelligence; DKEFS Stroop test switching scaled score; [a] Reading IQ measured on the NART unless the subject scored ≤15/50 on this test, in which case the Schonell Graded Word Reading Test IQ was used; [b] 1 PNFA patient did not perform these tasks (different subject for each); [c] 17 AD patients performed this task; [d] 8 controls performed this task

5.6.2. Peripheral hearing

Increasing age was associated with a significant increase in mean response time (detection threshold) at the three highest frequencies tested. Relative to the healthy control group, there was a significant difference ($p<0.05$) in mean detection thresholds for the AD group only at 0.5 kHz (4.1 dB) and 4 kHz (7.2 dB); these threshold elevations were small and unlikely to be clinically relevant, and there were no significant differences at any other frequency tested.

5.7. Results: Experimental tests

Table 5.3. Results for experimental tests in patient and control groups

		PNFA n=6		AD n=20		Control n=35	
		Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
Question comprehension[a]							
English accent	/40	31.4 (6.5)**†	24-40	38.0 (1.4)**	35-40	39.3 (0.2)	39-40
Foreign accent	/40	30.4 (6.7)**†	22-38	36.2 (2.2)**	31-39	38.6 (0.6)	38-40
Difference score: Foreign – English	/40	-1.0 (1.2)	-2.0, 1.0	-1.8 (2.2)	-8.0, 2.0	-1.3 (0.8)	-2.0, 0
Word verification[a, b]							
English	/24	22.6 (2.6)	18-24	22.6* (2.0)	16-24	23.7 (0.5)	22-24
American	/24	20.0 (3.5)*	14-22	21.6 (2.5)**	13-24	23.5 (0.7)	22-24
Australian	/24	19.6 (3.5)*	15-24	22.1 (2.5)	13-24	23.2 (0.8)	22-24
South African	/24	18.8 (4.7)	13-24	21.7 (1.9)	17-24	22.8 (1.0)	20-24
Difference score: Foreign[c] – English	/24	-2.4 (2.4)**‡	-5.0, 1.0	-0.8 (1.0)	-2.0, 1.3	-0.6 (0.7)	-2.3, 1.3
Phoneme discrimination							
Minimal pair word verification	/48	36.7 (13.7)	13-46	42.3 (5.5)**	24-47	46.7 (1.1)	44-48
Accent recognition							
English versus foreign (block 1)	/20	11.5 (4.1)**†	8-18	15.7 (3.0)**	9-20	18.8 (1.3)	14-20
English versus foreign (total)[d]	/80	60.7 (10.0)*	51-71	64.6 (8.9)**	45-75	75.1 (3.2)	64-79
British regions	/24	13.0 (4.8)**	7-19	14.9 (4.2)**	6-23	22.1 (2.5)	14-24
English regions	/24	15.7 (3.8)**	10-20	16.3 (2.5)**	12-21	21.3 (1.8)	18-24
Country knowledge							
Naming from description	/10	6.7 (2.1)**	4-10	7.9 (2.2)**	2-10	10.0 (0.2)	9-10
Map naming	/10	5.3 (1.8)**	3-8	6.0 (3.1)**	1-10	9.4 (0.9)	7-10
Map recognition	/10	7.5 (1.4)*	6-10	6.6 (3.4)**	0-10	9.9 (0.2)	9-10

Group differences significant after adjusting for background covariates (age, gender, and years of education) are displayed. [a] Results for N=5 PNFA subjects are shown (see text); [b] one AD patient declined to continue after three blocks and results were scaled to a score /24 for each accent for this subject; [c] mean score for all three foreign accents; [d] 3 PNFA subjects and 18 AD subjects were able to perform all 80 items on this test; *significantly worse than controls ($p<0.05$); ** significantly worse than controls ($p<0.01$); †significantly worse than AD group ($p < 0.05$); ‡ significantly worse than AD group ($p<0.01$)

5.7.1. Accent comprehension

Question comprehension

One patient with PNFA was unable to understand any questions on the first twenty items of the test (including items presented in the Southern Standard English condition); this patient was excluded from further analysis. Group results are presented in Table 5.3. All groups showed a reduction in mean scores for sentences presented in a foreign accent compared with Southern Standard English. Both patient groups showed a statistically significant reduction in score compared to controls in both foreign and Standard English accent conditions, and the PNFA group performed significantly worse than the AD group in both conditions. Foreign minus English difference scores did not differ significantly between any of the groups.

Word verification in healthy controls

Within the healthy control group, small but statistically significant differences in score were observed between word verification under foreign accents compared with the native English accent (see Table 5.3). Word verification scores were lower under all foreign accents compared with the English accent (mean difference in scores: American: -0.26 (95% CI: -0.46, -0.05); Australian: -0.54 (CI:-0.85, -0.24); South African: -0.89 (CI: -1.25, -0.53), all $p<0.0001$). Scores for the South African accent were also lower than for the more familiar American accent (difference: -0.64, (CI: -0.96, -0.29); $p<0.0001$) and for the Australian accent (difference: -0.34, (CI: -0.68, 0.01); $p<0.05$). Performance was weakly but non-significantly worse for the Australian accent than the American accent (difference: -0.29, (CI: 0.60, 0.03); $p=0.08$).

The reaction time analysis for items correctly identified showed strong evidence for prolonged mean reaction times (in seconds) for the three foreign accents compared to the English accent (American: 0.09, (CI: 0.06, 0.11); Australian: 0.11, (CI: 0.08, 0.15); South African: 0.06, (CI: 0.04, 0.09) all $p<0.0001$). Significant prolonged reaction time was also observed for Australian compared with South African accents (difference: 0.05 (CI: 0.01, 0.09) $p<0.05$). There was no evidence of differences between reaction times to other foreign accents ($p>0.05$).

Word verification across subject groups

Group results adjusted for age, gender and number of years of education, are presented in Table 5.3. One patient with PNFA performed at chance on all target accents, and was an outlier on

target and distractor items in the English accent (obtaining a score of 12/24 on English target items, and a score of 1/24 on English distractor items, where a chance score was 12 for each set). Due to the small patient group size, further analyses were conducted excluding this subject's data.

Compared to controls, the PNFA group showed a significant reduction in score for American and Australian accents (both $p<0.05$), but not South African (although there was a trend ($p=0.08$) to worse performance). The PNFA group demonstrated a significantly greater foreign minus English difference score compared to both control and AD groups (both $p<0.01$). The AD group performed significantly below controls for English and American accents ($p<0.05$), and there was a trend to worse performance for Australian ($p=0.07$) and South African accents ($p=0.09$), however the mean foreign accent minus English accent difference score was not significantly different to controls ($p>0.5$).

Phoneme discrimination

On this test of phonological processing ability, both patient groups showed a decrease in mean score compared to the control group, but differences only attained statistical significance in the AD group ($p<0.01$), the lack of statistical significance for the PNFA-control difference being attributable to a much large variability in PNFA scores. No significant difference was observed between the patient groups, although the mean score for the PNFA group was lower than for the AD group.

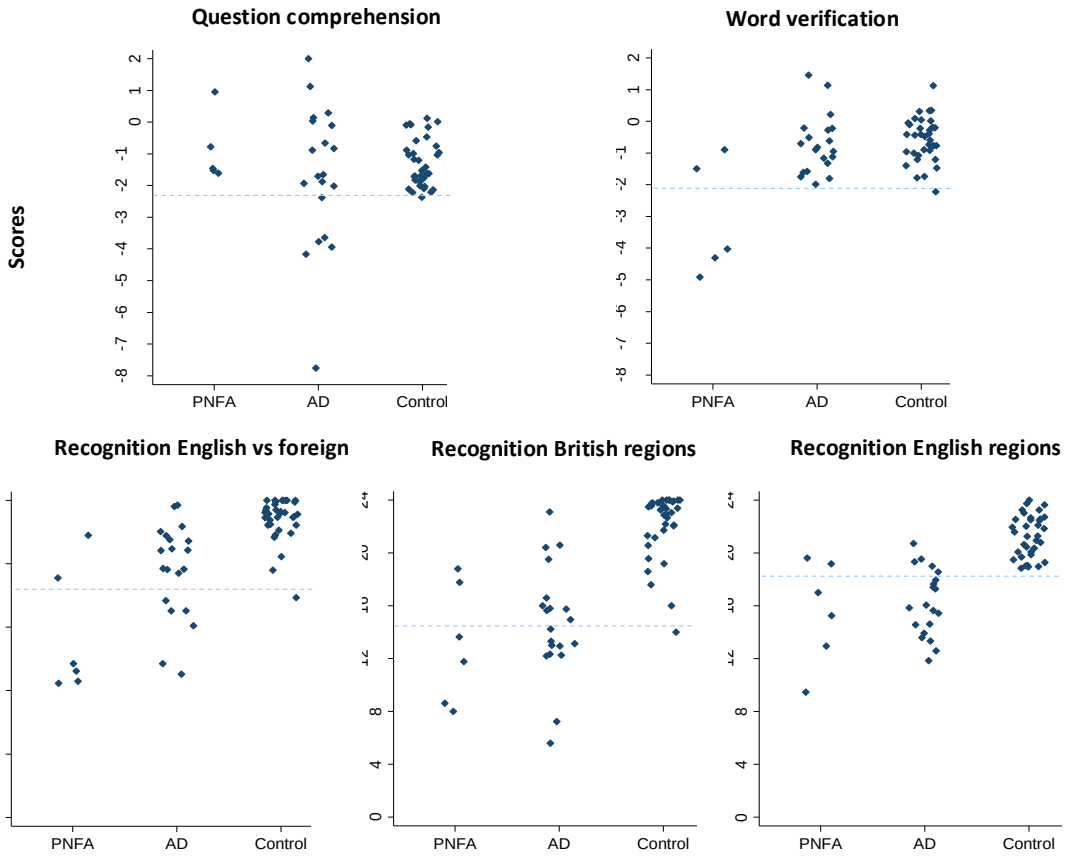
There was no evidence that difference score performance (Foreign minus English accent), for question comprehension and word verification) correlated with minimal pair phoneme discrimination in either patient group (see Appendix A.5.5). However as the PNFA group displayed a large decrease in score on the phoneme discrimination test, group differences in foreign minus English accent difference score were assessed after adjusting for performance on the phoneme discrimination task by adding phoneme discrimination test performance as an additional covariate to the regression model. The adjusted difference between PNFA subjects ($N=5$) and both the AD group ($p<0.05$) and the control group ($p<0.01$) respectively remained significant after adjusting for phoneme discrimination test performance.

Individual patient profiles

Raw subject data for Foreign minus English difference scores for accent comprehension tests are displayed in Figure 5.1. Individual subject performance was classed as impaired if below the 5th

percentile cut-off score for the healthy control group. On the question comprehension subtest 5/20 AD patients showed a large performance cost for foreign accents, falling below the 5th percentile of control values on the foreign minus English accent difference measure, while no PNFA patients (N=5, for whom data were available) fell below the 5th percentile. In contrast, on the word verification subtest 3/5 PNFA patients fell below the 5th percentile of control values on the foreign minus English accent difference measure, while no AD patients fell below the 5th percentile of control values.

Figure 5.1. Individual subject data for accent processing performance



Raw data for performance on experimental accent processing tasks are shown for individual subjects in the PNFA, AD and control groups. Dashed lines show 5th percentile cut-offs for each sub-test calculated from control data. Accent comprehension data are based on difference scores for question comprehension and word verification in Southern Standard English versus foreign accents (see text), where a negative score indicates increasing cost for presentation in a non-native accent. Scores (/20) for block 1 of the English-versus-foreign recognition test are shown; a score of 10 corresponds to chance performance on this test. For the regional British accent recognition test, a score of 6 corresponds to chance performance; for the regional English accent recognition test a score of 12 corresponds to chance performance.

5.7.2. Accent recognition

Foreign accents

Two patients with AD and three patients with PNFA performed near or at chance on the first block of this subtest, and therefore did not complete the full 80 item task. Group results adjusted for age, gender and education are presented in Table 5.3. Both patient groups performed significantly worse than control subjects ($p<0.01$) on both Block 1 of this test and the complete set of trials (which were performed in a reduced set of subjects as patients at chance on Block 1 were excluded) ($p<0.05$). The PNFA group performed below the AD group on Block 1 of this test ($p<0.05$) and for the complete set of trials, although the latter difference was not statistically significant.

Regional accents

A similar profile was found for recognition of British and English regional accents: both patient groups performed significantly worse than controls ($p<0.01$) for recognition of British regional accents, however there was no statistically significant performance difference between the two disease groups.

Country knowledge

The AD and PNFA groups performed significantly worse than the control group on all country recognition subtests ($p<0.05$). There was no significant performance difference between the two disease groups. There was little evidence of correlation between performance on accent recognition tests and tests of country knowledge in the AD group (see Appendix A.5.5); only performance on the British accent recognition test showed a positive correlation with country naming from description ($r=0.52$, $p<0.05$) and no other significant correlations were observed. As both patient groups were significantly impaired on tests of country knowledge, differences between the subject groups on accent recognition subtest performance were additionally analysed adjusting for performance on each test of country recognition. Differences between patient groups and controls on accent recognition tests remained significant after adjusting for performance on these tests ($p<0.05$).

Individual patient profiles

Raw subject data for accent recognition tests are displayed in Figure 5.1. A high proportion of patients in both disease groups (4/6 PNFA, 17/20 AD) performed below the 5th percentile control

score on at least one accent recognition subtest. In the PNFA group, the same 4/6 patients fell into the impaired range on all three subtests. In contrast, in the AD group, 6/20 subjects were impaired on block 1 of the foreign versus English accent subtest, 10/20 patients on the regional British accents subtest and 14/20 patients on the regional English accents subtest; only 4/20 patients were impaired on all three subtests.

5.8. Correlations of accent processing performance with neuropsychological measures and tests of apperceptive and semantic voice processing

Correlations between tests of accent processing and both background neuropsychological and voice processing in the AD group are displayed in Appendix A.5.6. All three tests of accent recognition correlated with two measures of voice recognition: familiarity and naming tests, in addition British and English regional accents tests correlated with identification and cross-modal matching. English versus foreign accent recognition test performance significantly positively correlated with difficult speaker discrimination ($p<0.05$), whereas Foreign minus English word verification difference score showed a significant negative correlation with easy speaker discrimination ($p<0.05$).

Several tests of accent recognition performance significantly positively correlated with tests of episodic memory: British regional accent recognition performance correlated with recognition memory for words and England versus foreign accent recognition correlated with recognition memory for faces ($p<0.05$). Foreign minus English question comprehension difference score showed a significant positive correlation with digit span forwards and semantic comprehension on the BPVS, British regional accent recognition also significantly positively correlated with the BPVS as well as performance IQ (all $p<0.05$).

5.9. Neuroanatomical data

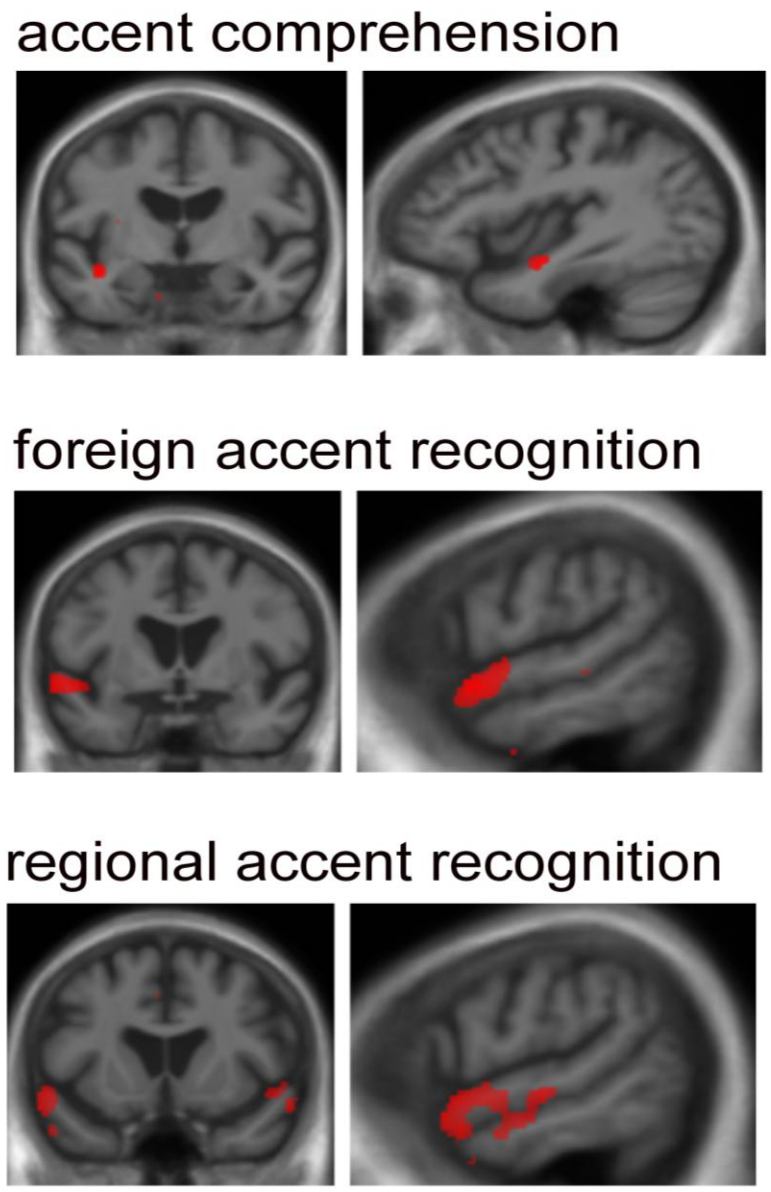
Results of the neuroanatomical data are summarised in Table 5.4 and statistical parametric maps are shown in Figure 5.2.

Table 5.4. VBM data: neuroanatomical associations of experimental test performance in the patient groups

	Side	Region	Z score	Cluster size (voxels)	MNI Coordinates (mm)		
WITHIN-AD GROUP (n = 17)							
Accent comprehension							
Difference score: Foreign – English questions	Left	Anterior STG	4.74	125	-42	-8	-18
Accent recognition							
British regions	Right	Anterior STG	4.55	214	50	16	-11
COMBINED GROUPS (n = 21)							
Accent comprehension							
Difference score: Foreign – English questions [a]	Left	Anterior STG	4.38	78	-42	-8	-18
Accent recognition							
English versus foreign (block 1)	Left	Anterior STG/STS	4.36	765	-61	8	-12
British regions	Left	Anterior STS/STG/MTG	4.86	1704	-64	-3	-11
	Right	Anterior STG	4.39	162	50	16	-11

Areas listed are based on local maxima exceeding a voxel-wise significance threshold after FWE-correction over the prespecified small volume of interest. All clusters of size >10 voxels are shown. Z scores refer to the local maxima within these regions. MNI, Montreal Neurological Institute; MTG, middle temporal gyrus; [a] results for N=20 patients are shown as one PNFA subject was unable to perform this test (see text).

Figure 5.2. Statistical parametric maps of grey matter volume associated with accent processing performance



SPM shows grey matter associations of experimental test performance across the AD and PNFA groups (see also Table 5.4): top, difference score on question comprehension under foreign versus standard English accents (accent comprehension); middle, recognition of foreign vs. standard English accents (foreign accent recognition); bottom, recognition of regional British accents (regional accent recognition). SPMs are presented on sections of the mean normalised T1-weighted structural brain image in DARTEL space. Coronal (left) and sagittal (right) sections are shown. The sagittal sections are derived from the left hemisphere and the left hemisphere is shown on the left in the coronal sections. All SPMs 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 (see Table 4); SPMs are thresholded at $p<0.001$ uncorrected for display purposes.

Combined data across groups

The analysis combining disease groups (n = 21) revealed no significant grey matter associations for group-performance interactions for any of the experimental subtests, over either the whole brain or the pre-specified temporal lobe volumes of interest. There were no significant grey matter associations of experimental test performance across groups after whole brain correction. Restricting analyses to the pre-specified temporal lobe volumes of interest, several associations were identified (all $p<0.05$ after FWE correction over the small volume). The foreign minus English accent difference score for the question comprehension subtest (n=20, as one PNFA subject was unable to perform this test) was positively associated with grey matter in left anterior STG. Performance on the foreign accents recognition test was also positively associated with grey matter volume in left anterior STG and STS; while performance on the regional British accents test was positively associated with grey matter volume in left anterior STS / STG/ MTG, and right anterior STG. No significant grey matter associations were identified for performance on other experimental subtests.

Within-group data

The AD group (n = 17) showed no significant grey matter associations of experimental test performance after correction for multiple comparisons over the whole brain volume; however, restricting analyses to the pre-specified temporal lobe volumes of interest, the foreign minus English accent difference score (question comprehension subtest) was positively associated with grey matter in left anterior STG, while performance on the regional British accents subtest was positively associated with grey matter volume in the right anterior STG (all $p<0.05$ after FWE correction over the small volume). The AD group-only cluster maxima for these experimental tests were identical to those for the combined group analyses (see Table 5.4), and therefore the statistical parametric maps are not displayed. The PNFA group (n=4) showed no significant grey matter associations of experimental test performance.

5.10. Discussion

In this study impairments of non-native accent comprehension and recognition were demonstrated in patients with two canonical dementia syndromes, AD and PNFA. Both patient groups showed impaired recognition of foreign and regional accents; at the individual subject level patients with PNFA showed a more consistent pattern of impairment over different (foreign and regional) levels of accent recognition. The PNFA group also showed reduced comprehension of words spoken in foreign accents compared with a Southern Standard English accent. Individual

subject data suggested dissociable patterns of impairment under foreign accents: while patients with PNFA frequently showed a perceptual cost for comprehension of accented words (but not sentences), patients with AD frequently showed a perceptual cost for comprehension of accented sentences (but not single words). These deficits were not clearly attributable to a general phonological or geographical semantic impairment.

Information about accent processing in neurodegenerative disease is very limited. However, the present findings add to previous evidence for impairments of other aspects of complex auditory pattern processing in these diseases. In particular, deficits in the perception and comprehension of prosody have been demonstrated in AD (Allender *et al.*1989; Horley *et al.*2010; Roberts *et al.*1996; Taler *et al.*2008; Testa *et al.*2001) and PNFA (Rohrer *et al.*2010). As another example of a meta-linguistic vocal signal with segmental, suprasegmental and semantic dimensions, prosody is expected to engage brain mechanisms similar to those involved in accent processing. However, previous studies of nonverbal sound processing in AD and PNFA suggest that these diseases may affect distinct components of vocal signal analysis: whereas AD is predominantly associated with deficits of sound pattern analysis under non-canonical listening conditions (Gates, Beiser, Rees *et al.* 2002; Goll *et al.*2012), PNFA is predominantly associated with conjoint deficits of timbre and auditory semantic processing suggesting a more fundamental deficit in the encoding of auditory object properties (Goll *et al.*2010; Goll *et al.*2012). These core deficits might contribute to the dissimilar patterns of accent comprehension impairment (perceptual cost) shown by individual patients with AD versus PNFA: whereas comprehension of questions is likely to depend on tracking extended auditory patterns, comprehension of monosyllables is more likely to depend on accurate encoding of individual sound objects (here, spoken phonemes). A primary perceptual deficit might lead to degraded representation of accent characteristics and consequently reduced recognition of those accents, or conversely, impaired accent knowledge might damage 'top-down' mechanisms that normally act to disambiguate the effects of perceptual distortion, as has been suggested in previous work. It has been hypothesized that adaptation to unfamiliar accents engages top-down lexically driven categorisation mechanisms (Bradlow & Bent 2008; Norris *et al.*2003); such dynamic mechanisms could plausibly be degraded in neurodegenerative disease and could be assessed in future work.

Although tests of potentially relevant vocal perceptual processes such as prosodic contour discrimination (Rohrer *et al.*2010) were not assessed in this study, the present data do not suggest a clear, consistent perceptual defect across the disease groups. English versus Foreign accent

recognition performance positively correlated with performance on a test of voice perception: difficult speaker discrimination (tested in the AD group only). It is possible that this may represent a shared apperceptive level of processing as the English versus Foreign accent test relied on an ability to distinguish a 'canonical' from a non-canonical accent.

The results of this study suggest that additional semantic-level deficits may also play a role in impaired recognition of non-native accents in AD and PNFA. It is plausible a priori that the semantic processing of accents might be aligned with other geographically-organised concepts (Crutch *et al.*2003; Crutch *et al.*2010; della Rocchetta *et al.*1998); the present data provide only limited support for correlated performance on accent and geographical knowledge (though this correlation was assessed in the AD group only) and suggest that accent recognition is not merely subsumed by brain mechanisms of geographical semantic processing. A consistent pattern of correlation (again tested in the AD group only) was found between tests of accent recognition and familiar voice recognition (results from Study 2) suggesting that impairments on other semantic vocal tasks accompany accent recognition deficits, at least in AD.

The present neuroanatomical findings corroborate these behavioural profiles. Both in the AD group alone and across groups, a measure of accent comprehension was positively associated with grey matter volume in left anterior STG, however interpretation of this association should be cautious in the absence of a clear overall behavioural cost relative to healthy controls. Recognition of foreign and non-native regional accents was positively associated with grey matter volume in a more anterior cortical region in left anterior STG / STS, while more fine-grained recognition of regional accents was additionally associated with grey matter volume in right anterior STG. It is noteworthy that cortical associations were found within temporal lobe areas close to cortical associations of voice recognition identified in the previous study (Chapter 4) but somewhat more anterior than those previously implicated in certain other aspects of nonverbal perceptual analysis (Norris *et al.*2003; Rohrer *et al.*2010). This might reflect shared mechanisms for processing the meaning of accents and other dimensions of the vocal signal: the processing of accents may depend on brain mechanisms analogous to those mediating speech intelligibility under other forms of perceptual distortion (Binder *et al.*2000; Bishop *et al.*2009; Friederici, Kotz, Scott *et al.* 2010; Leff, Iverson, Schofield *et al.* 2009; Scott, Blank, Rosen *et al.* 2000; Scott *et al.*2006; Zatorre *et al.*1992). Indeed, accented speech could be viewed as an 'ecological' example of degraded speech, representing an extreme form of the phonological–phonetic variation exhibited by individual speakers even within the spectrum of a native accent (Best *et al.*2001;

Best, McRoberts, & Sithole 1988; Clarke *et al.*2004; Flege, Munro, & MacKay 1995; Floccia *et al.*2006; Iverson & Kuhl 2000; Iverson, Kuhl, Akahane-Yamada *et al.* 2003; Norris *et al.*2003; Nygaard *et al.*1998). Comprehension of accented speech may involve assimilation of accented phonemes into categories used for native speech, for example involving matching to stored prelexical templates (Best *et al.*2001; Best *et al.*1988; Clarke *et al.*2004; Flege *et al.*1995; Floccia *et al.*2006; Iverson *et al.*2000; Nathan *et al.*1998). Tolerance to phonetic variation is likely to be established via exposure to many individual speakers with different accents (Adank *et al.*2009; Bradlow *et al.*2008; Clopper *et al.*2008; Clopper *et al.*2004b; Floccia *et al.*2006). The putative template matching algorithm has been shown to be inflexible in infants (Nathan *et al.*1998; Schmale *et al.*2009) and may be disrupted in neurodegenerative disease.

While stored representations of single phonemes are likely to be instantiated in posterior superior temporal cortices (Chang *et al.*2010; Jancke *et al.*2002; Rauschecker & Scott 2009; Turkeltaub *et al.*2010), decoding of extended utterances such as questions posed in a foreign accent is likely to require tracking of auditory information streams over longer time periods, a function previously localised to more anterior temporal cortices (Friederici, Meyer, & von Cramon 2000; Humphries, Willard, Buchsbaum *et al.* 2001; Meyer *et al.*2002; Scott *et al.*2000). Comprehension of spoken sentences in AD may also have been influenced by working memory capacity: in this study correlation was demonstrated with digit span (Table 5.6.), and previous work in AD has associated deficits of auditory stream analysis with working memory impairments (Goll *et al.*2012).

A complementary interpretation of the present data would hold that accent comprehension depends on stored knowledge about accent properties that also supports accent recognition: in particular potentially access to representations of linguistic features in the left ATL, supporting the preliminary model of accent processing proposed in this thesis (described in Section 1.7.3). Models of speech comprehension and voice recognition (Belin *et al.*2004; Belin *et al.*2000; Scott *et al.*2006) assign to the anterior STG / STS a key role later in the cortical processing hierarchy for auditory "what" information. Although our analyses are likely to have been underpowered to differentiate laterality effects between tasks, as predicted in the model of voice processing for this thesis recognition of accents was associated with the ATL bilaterally.

Neuropsychological and neuroanatomical overlap between accent recognition and voice recognition suggests that semantic representations of familiar accents as well as familiar voices

may be stored in the right ATL in line with the view that this area represents knowledge about nonverbal, socially relevant vocal signals (Olson *et al.*2007; Omar *et al.*2010; Omar *et al.*2010; Rosen *et al.*2006). The results suggest there may be at least partial differentiation between these representations with voice and cross-modal person recognition in inferior parts of the ATL and associations with accent recognition in more superior parts of the right ATL, as predicted in the model (Section 1.7.3). The results further suggest that involvement of the right ATL in voice processing more broadly is not likely to be solely dependent on representation of timbre, as accent recognition is likely to be more dependent on prosodic cues such as pitch contour and stress cues. Rather, the right ATL may be fundamentally concerned with processing representations of complex spectrotemporal configurations embodied in voices and accents, and in particular, associating these with meaning. This would be in line with auditory object processing models proposing that there is close correspondence between apperceptive and semantic mechanisms. Investigation of how the processing of non-native accents relates to the processing of other vocal properties is needed to further develop models of voice processing.

The findings suggest that impairments of accent processing may constitute signatures of neurodegenerative diseases and not merely amplification of an effect already present in the normal brain. Healthy control subjects here showed a performance profile across non-native accents that could reflect past exposure and familiarity with those accents (Adank *et al.*2009; Clopper *et al.*2008; Clopper *et al.*2004a) or alternatively, the relative perceptual similarity of the accents chosen here to Southern Standard English (see Supplementary Material on-line), in line with previous suggestions (Clarke *et al.*2004; Flege *et al.*1995; Floccia *et al.*2006; Iverson *et al.*2000; Norris *et al.*2003). Processing of accents is potentially a test case with much broader implications for understanding how the brain encompasses perceptual variation in behaviourally relevant, semantically laden stimuli and how neurodegenerative diseases damage the distributed cortical networks that are presumed to support such processing.

This study has several limitations and suggests a number of directions for future work. Accent processing here was assessed in relation to a limited number of other neuropsychological functions: a more complete understanding of the deficits identified here would require a more detailed investigation of accent processing in parallel with other kinds of complex nonverbal sound processing and a more fine-grained analysis of potentially relevant perceptual and linguistic mechanisms. This study aimed to address a broad range of accent processing functions (aspects of accent comprehension and recognition) for accents that were likely to be familiar to

our subject population and using various relevant response procedures (sentence comprehension, word verification and forced-choice responses): future work should analyse the component processes in more detail and compare these processes more directly using uniform test procedures. It will be important to assess performance in relation to the specific perceptual characteristics that define particular accents, a key issue in attempting to generalise findings across populations with very different accent exposures.

A further dimension is the potential interaction between altered accent perception and distorted production of the patient's own native accent, as illustrated most dramatically in the so-called 'foreign accent syndrome' (Hall *et al.*2003; Kurowski *et al.*1996; Luzzi *et al.*2008; Van Borsel *et al.*2005): this would entail a parallel acoustic analysis of patients' spoken output. Even if temporal lobe areas are critical for accent processing, such processing is likely to be mediated by distributed brain networks extending beyond the temporal lobes (Adank *et al.*2012; Berman *et al.*2003; Burton, Small, & Blumstein 2000; Peschke, Ziegler, Eisenberger *et al.* 2012). As in Study 2, grey matter associations were only significant in ROI analyses, and associations were not found with inferior frontal and parietal cortical regions hypothesised to be recruited in accent processing tasks, VBM analyses may have been underpowered to detect associations across the whole brain, or alternatively these regions may not have been critical to impaired task performance in PNFA and AD. A more complete picture of these mechanisms will require complementary functional and connectivity-based imaging techniques, in line with the emerging concept of neurodegenerative diseases as 'network-opathies' (Buckner, Sepulcre, Talukdar *et al.* 2009; Seeley *et al.*2009; Sonty *et al.*2007; Zhou *et al.*2010).

A further limitation is that the patient cohorts here were relatively small and this limitation is likely to be particularly relevant to intrinsically heterogeneous syndromes such as PNFA; tests for disease-performance interactions here were likely under-powered, and the PNFA group did not contribute substantially to the combined-group results. There is a need to address these issues in larger patient cohorts, in other neurodegenerative diseases and longitudinally, in order to establish how accent processing relates to the development of other cognitive deficits and the specificity of deficits for particular neurodegenerative pathologies.

6. Conclusions

The work presented in this thesis has addressed the neuropsychological and neuroanatomical underpinnings of voice processing impairments in neurodegenerative disease. The present experimental evidence has implications for our understanding of the pattern of nonverbal auditory impairments in FTLD and AD, and for the architecture of voice processing in the healthy brain. In this chapter a summary of the findings of this thesis and directions for future work are presented.

1. Voice processing deficits are significant in neurodegenerative disease

Detailed studies of voice processing in degenerative disease are limited, however the findings of this thesis suggest that vocal processing impairments are likely to be prevalent in both AD and FTLD syndromes, in keeping with predictions that the brunt of tissue damage in these diseases involves a network of temporal, frontal and parietal cortical regions which are likely to contain mechanisms integral for voice and accent analysis (Adank *et al.*2009; Belin *et al.*2004; Berman *et al.*2003; Van Lancker *et al.*1988; Van Lancker *et al.*1989; von Kriegstein *et al.*2004). Further investigations will be required in larger patient cohorts, however the current findings suggest that different dementia syndromes lead to distinct but partly overlapping profiles of voice processing impairment.

2. Associative deficits predominate in these syndromes

Semantic deficits of voice processing (recognition of voices or accents) were found in all the dementia syndromes investigated in this thesis: SD, bvFTD, PNFA and AD. The selective nature of the breakdown of conceptual knowledge in right and left-sided tvFTLD (primarily consisting of cases that fulfilled clinical criteria for SD) was a critical test case to investigate the nature of semantic voice processes. In common with deficits of recognition of other nonverbal auditory objects (Bozeat *et al.*2000; Goll *et al.*2010; Omar *et al.*2010; Omar *et al.*2011), severe associative deficits were generally observed in the presence of relatively preserved perceptual and apperceptive vocal task performance, and alongside multimodal deficits of person knowledge. VBM correlates further implicated the anterior and inferior temporal lobes in the pan-modal semantic analysis of sensory objects. The neuroimaging evidence presented in this thesis suggests dedicated brain regions representing modality-specific information within a distributed bi-temporal network instantiate mechanisms for processing multiple aspects of knowledge. The right

ATL was implicated particularly for aspects of nonverbal knowledge (voice recognition (Study 2) and accent recognition (Study 3)), while the left ATL may be more important in verbally mediated knowledge (single word comprehension on the BPVS (Study 2) and the intelligibility of accented speech (Study 3)).

In previous work, deficits of voice processing in AD and PNFA, in particular studies of prosodic discrimination, were attributed to apperceptive auditory object impairments associated with posterior temporo-parietal atrophy (Goll *et al.*2010; Rohrer *et al.*2010). The present data however suggest that vocal semantic deficits associated with atrophy to the anterior and superior temporal lobes are prevalent in both syndromes. The proximity of neuroanatomical associations in the right ATL between accent and voice semantic processing tasks, as well as strong evidence for correlation between accent recognition and voice recognition task performance in the AD group suggest overlapping impairments between two vocal tasks. In AD these deficits are likely to be part of a broader profile of nonverbal semantic deficits associated with ATL atrophy in this disease (Greene *et al.*1996; Hodges *et al.*1992; Lambon Ralph *et al.*2003; Perry *et al.*2000), whereas in PNFA, receptive vocal processing deficits such as perception of "non-canonical" foreign-accented phonemes may provide a point of convergence between receptive and characteristic expressive speech impairments such as abnormalities of accent or prosody (Rohrer *et al.*2010).

3. Apperceptive and semantic vocal processing stages may interact

In parallel with processing of visual information in the face modality, hierarchical cognitive models of voice processing propose that apperceptive processes precede semantic mechanisms of analysis (Belin *et al.*2004; Bruce & Valentine 1985; Ellis *et al.*1997). The results of this thesis present a more complex picture. Study 2 found that deficits of speaker discrimination were not consistently related to semantic task performance in AD, and although speaker discrimination correlated with voice familiarity, the direction of the influence was not established, and could involve reciprocal interaction (Goll *et al.*2010). BvFTD case QR (Study 1) also raised the possibility that selective associative phonagnosia could arise as a result of an abnormal interaction between complex perceptual representations (for example timbre, articulation or prosody) and subsequent semantic mechanisms. Close association and interaction between perceptual and semantic stages of nonverbal processing has been suggested in models of person identification (Lucchelli *et al.*2008) and auditory object models in which increasingly cross-modal information and semantic processes are integrated as increasingly complex

spectrotemporal representations are formed (Goll *et al.*2010; Griffiths *et al.*2007; Lewis *et al.*2009).

Feedback from cross-modal or semantic processes to perceptual analyses may be particularly important under non-standard listening conditions (e.g., identification of voices over the phone or when singing: (Benzagmout *et al.*2008; Garrido *et al.*2009)) when the auditory system is confronted with ambiguous or novel auditory information such as when confronted with a speaker with a novel foreign accent. Under such conditions, dynamic updating and generalization of conceptual representations may be necessary, for example updating of spectrotemporal representations of a linguistic unit and stored representations of particular accent features. The functional underpinnings of deficits of accent recognition and vocal identity processing associated with regions in the right ATL were not evaluated in this thesis, but may have resulted from defective perceptual differentiation amongst closely related auditory entities and/or fine-grained analysis at associative (semantic) processing stages. To investigate the nature of deficits at the ATLs further neuropsychological and neuroimaging paradigms are needed, manipulating perceptual and semantic demands.

4. Vocal apperceptive processing utilises auditory and non-auditory cognitive mechanisms

Neuroimaging work suggests voice-specific mechanisms for perceptual analysis of voices exist in bilateral regions of the upper STS. Study 2 found that vocal apperceptive deficits in AD subjects were associated with regions of the right inferior parietal cortex, raising the possibility that vocal apperceptive mechanisms may depend on a distributed temporo-parietal network, involving both analysis of critical auditory features as well as other cognitive mechanisms, such as auditory working memory. This is concordant with deficits of nonvocal auditory object analysis, which are usually accompanied by impaired lower level object property processing and/or other non-auditory cognitive processing deficits (Goll *et al.*2012; Griffiths, Rees, Witton *et al.* 1997; Saygin *et al.*2010).

In the model of voice processing presented in this thesis, auditory working memory mechanisms are also proposed to be integral for accent processing. However differences between perceptual and apperceptive processing of voices and accents are hypothesised in the model, in particular accented speech presents listeners with a "non-canonical view" of a phoneme and may therefore utilise apperceptive mechanisms for another category of auditory objects: speech sounds. Voice specific apperceptive deficits are rarely detected clinically, hence systematic investigation of

voice processing in brain-lesioned cases presenting with other auditory deficits (such as dystimbria, aprosodia or word deafness, where voice processing deficits have been implicated in previous work (Mazzucchi *et al.*1982; Peretz *et al.*1994)) and in cases with auditory working memory impairment is needed to establish the critical neuro-cognitive mechanisms for voice and accent perception.

Speaker discrimination is a neuropsychological paradigm developed by analogy with tests of facial and visual apperception (Benton *et al.*1989; Warrington *et al.*1986) and it is possible that comparing different "views" of unknown speakers addresses voice processing mechanisms that operate in parallel to voice recognition; for example recognition of familiar speakers may be less dependent on comparison of timbre between consecutive voices in working memory, and more reliant on fine-grained matching of spectrotemporal information to stored complex representations. A neuropsychological apperceptive task which does not require comparison of vocal features across a delay needs to be developed, for example a task which requires subjects to decide whether a sound is a "voice or not a voice" (a similar idea was utilised in a functional imaging paradigm previously (Belizaire *et al.*2007)) may parallel apperceptive tests of environmental sound processing in which subjects decide whether environmental sounds are real or unreal (Goll *et al.*2010), analagous to the Object Decision Task in the visual modality. Alternatively a test in which subjects must decide how old an unfamiliar voice is, by analogy with the De Renzi test of facial apperception (De Renzi, Faglioni, Grossi *et al.* 1991; De, Bonacini, & Faglioni 1989) could be developed. Further neuroimaging and neuropsychological investigation of vocal perceptual and apperceptive processes is needed to establish the nature and extent of analogies between auditory and visual modalities.

5. Familiar voice identification segregates from recognition of other auditory objects in degenerative disease

The data presented in Study 1 found that recognition of environmental sounds was not impaired with phonagnosia in neurodegenerative disease, supporting the results from previous work showing dissociation between vocal associative processes and this class of auditory object (Assal *et al.*1981; Garrido *et al.*2009; Neuner *et al.*2000; Peretz *et al.*1994). This result contrasts with the results of previous work showing impairments of environmental sound recognition in SD (Bozeat *et al.*2000; Goll *et al.*2010). Both cases in Study 1 however performed in the impaired range on musical instrument identification; music is another specialized category of auditory objects involving highly differentiated auditory entities (musical instruments and melodies) and musical

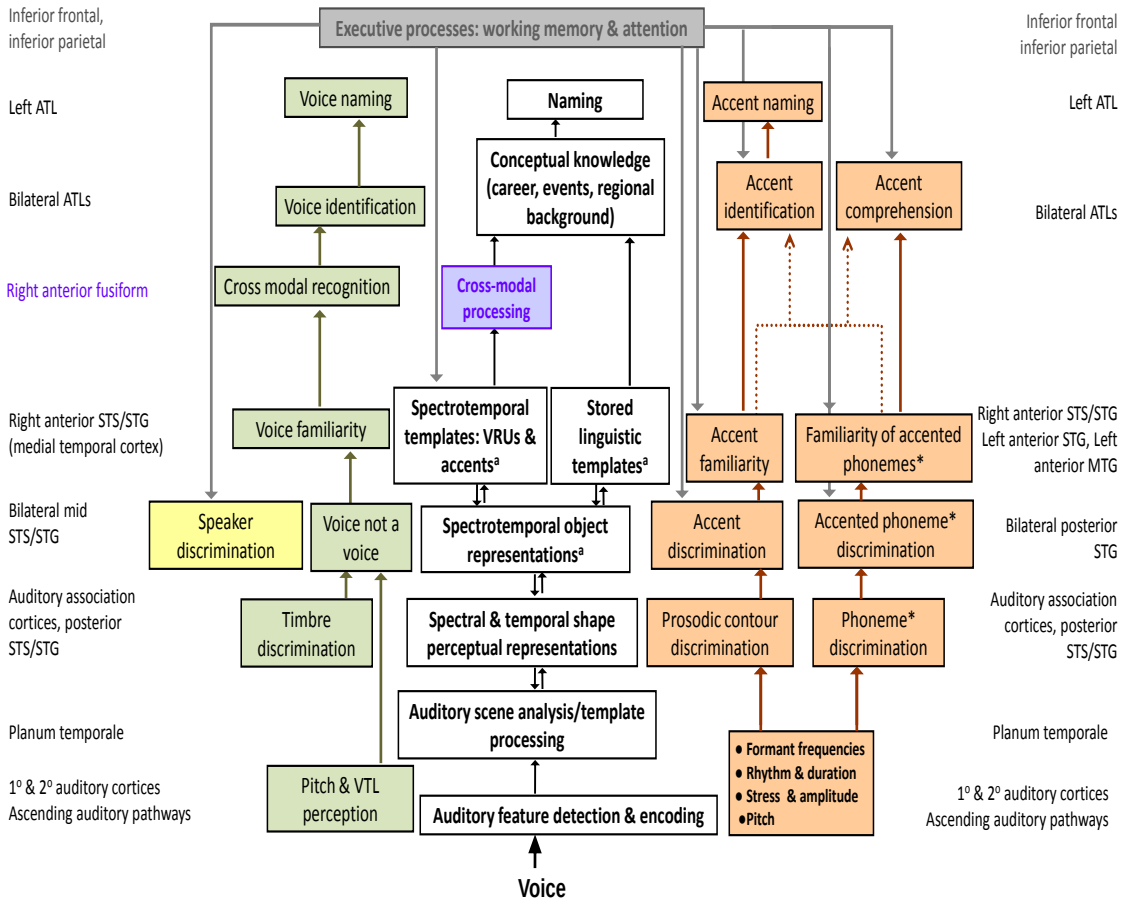
instrument recognition, like voice processing is highly dependent on analysis of timbre. A neuroanatomical profile either affecting mid temporal cortical regions and disrupting mechanisms of timbre analysis, but sparing posterior regions of the STS critical for analysis of lower level auditory perceptual features sufficient for environmental sound discrimination could explain this result (Goll *et al.*2010; Lewis *et al.*2009). Investigation of auditory object processing in larger groups of phonagnosic individuals is needed; musical auditory objects may be valuable for matching processing demands to voices (for example timbre processing between instruments and individual voices, and melodic processing between melodies and accented voices). Finding another auditory category which is as highly differentiated as familiar voices may be difficult in listeners without auditory expertise; regional accents may offer a less highly differentiated class of vocal semantic object to compare to other auditory object categories which do not require expertise (such as tool or animal sounds) (Goll *et al.*2010; Goll *et al.*2010; Lewis *et al.*2005; Staeren *et al.*2009).

The data presented in Study 1 also demonstrated that phonagnosia can spare recognition of vocal emotions in FTLD, parallels the results exhibited by developmental case KH (Garrido *et al.*2009), and corresponds to dissociations between identity and emotion pathways in the face modality. However, right-lateralised regions of the amygdala associated with familiar voice recognition in Study 3 raised the possibility that there is greater overlap between emotion and identity processing pathways than proposed by Belin (see Section 1.7.2) (Campanella *et al.*2007; von Kriegstein *et al.*2004). Interaction with affective processing may occur at the level of processing of object properties (such as timbre or intonational qualities); such interactions have been tested in music: melodies have been shown to be adept at conveying emotions and abstract feeling states (Gosselin, Peretz, Hasboun *et al.* 2011; Gosselin, Peretz, Johnsen *et al.* 2007; Peretz & Zatorre 2005), musical instrument timbre has been found to communicate emotional tone (Hailstone *et al.*2009). The emotional qualities of voices could be tested empirically. Recognition of personally familiar voices may also be more dependent on affective associations than recognition of public figures, and could for example involving arousal via the amygdala as has been suggested underpins covert face recognition (Schweinberger & Burton 2003). Further work is needed to establish whether personally familiar voices are affected similarly to recognition of public figures.

6. Updated model of voice processing in this Thesis

On the basis of the results from thesis, the model of voice processing outlined in Chapter 1 can be updated. This updated model is displayed in Figure 6.1. Hypotheses arising from the model are discussed below and directions for future work are suggested.

Figure 6.1 Updated model for this Thesis



Developed from Figure 1.5 on the basis of neuropsychological and neuroanatomical data from the studies in this thesis. Parallel processing pathways for accent processing (in orange), voice identity processing (in green) and speaker discrimination (in yellow) are indicated with their component cognitive operations. Candidate anatomical substrates for these operations are displayed on the relevant side of the diagram, and although they are displayed as discrete "nodes", it is likely that areas cooperate as networks. Accented processing is proposed to recruit linguistic processing mechanisms, mechanisms are indicated by analysis of phonemes*, however the linguistic mechanisms may operate at the level of speech sounds: consonants and vowels, phonemes, phonological units and/or at the level of words: these processes were not specifically investigated in this thesis. [a]Apperceptive voice processes are indicated by two stages, and although are poorly understood and were not specifically investigated in this thesis, in this model generic auditory object analysis has been segregated from more individuated analysis of previously encountered voices (VRUs) and linguistic templates for the purposes of distinguishing the process of familiarity. Coloured arrows demonstrate the primary information transfer pathway and indicate the primary direction of communication between stages in the hierarchy. Although the main pathways of information transfer are displayed, connections may be reciprocal and some lateral interaction between parallel pathways is likely. Arrows linking template processing and spectral and temporal shape representations are shown as bidirectional to emphasise the dynamic updating of these templates via the interaction between incoming information and stored representations, in line with Goll and colleagues' model of auditory object processing (Goll *et al.*2010).

On the basis of neuroimaging results of Studies 2 and 3 in which the right ATL was implicated in accent and voice recognition performance, it is hypothesised that the right superior anterior temporal cortex contains complex spectrotemporal representations or templates of both familiar voices (VRUs) and of regional and foreign accents. It is therefore predicted that damage to the right anterior superior temporal lobe could result in impairments of voice and accent recognition tasks. Regions of the left anterior superior temporal lobe (STS/STG) and anterior mid temporal gyrus predicted to be involved in higher level linguistic analysis were associated with both accent recognition and comprehension tasks in Study 3, and hence it is predicted that damage to the left ATL would severely affect accent comprehension and recognition performance. Associations with bilateral regions of the temporal pole were found across vocal semantic tasks in Study 2, regions involved in representations of amodal or multimodal semantic knowledge (Humphreys *et al.*1988; Lambon Ralph *et al.*2008; Snowden *et al.*2004; Warrington 1975), therefore it is predicted that selective damage to the left ATL would impair identification and naming tasks but may not impair voice familiarity or nonverbal cross-modal recognition, such as voice- face matching. These predictions could be tested systematically in unilateral temporal lesion cases as well as patients with right or left predominant temporal lobe atrophy.

Study 2 found consistent associations across person recognition tasks with grey matter in the right anterior and inferior temporal lobe. Familiarity decisions across modalities were associated with the right anterior fusiform gyrus, in line with models of person recognition which hypothesise that multimodal familiarity decisions do not require access to conceptual knowledge (Belin *et al.*2004; Bruce *et al.*1986; Burton *et al.*1990), represented at the poles. It is possible that the right anterior fusiform gyrus is the neural locus for the PIN, in which damage here would result in multimodal deficits of person recognition, but not deficits of other semantic vocal tasks such as accent recognition. A neural mechanism for multimodal familiarity decisions does not easily explain neuropsychologial evidence in Studies 1 and 2 for dissociations between performance on familiarity tasks across modalities. Notably, dissociation was found between verbal and nonverbal modalities, corroborating results found in previous work (Gainotti 2007a; Gainotti 2007b; Snowden *et al.*2004). Neuroanatomical segregation between modalities however was not seen in VBM analyses, which may have been underpowered to detect modality specific effects.

The model in this thesis is based on a recent model (Gainotti 2011), which argues against the existence of a PIN based on evidence for modality specific deficits at familiarity tasks (Gainotti 2007a; Gainotti 2007b; Gainotti 2011), proposing that familiarity decisions occur in modality

independent recognition units in line with the original Bruce and Young model (Bruce *et al.*1986). Gainotti's model is complementary to a model proposing direct and reciprocal connectivity between modality-specific recognition units (von Kriegstein *et al.*2006), based on work showing functional connectivity between STS and fusiform during familiar speaker recognition (von Kriegstein *et al.*2006; von Kriegstein *et al.*2005). The model for this thesis predicts that deficits of voice familiarity and associative phonagnosia (in Study 1) can result from selective damage to superior regions of the right ATL. It is proposed that damage to the anterior fusiform would impair voice cross-modal recognition or identification tasks.These hypotheses could be tested in patients with lesions or focal atrophy affecting the right anterior fusiform, in line with previous study of voice recognition in prosopagnosia (von Kriegstein *et al.*2006).

In the model familiarity is also hypothesised to be mediated by attention or executive resources, particularly under challenging or non-standard listening conditions (Benzagmout *et al.*2008; Garrido *et al.*2009). Further exploration of modality specific differences in neural networks is needed, and manipulations of both the modality of presentation and listening conditions could be readily explored using fMRI in patients as well as healthy subjects. In particular fMRI would allow the functional connectivity of voice networks to be explored (for example von Kriegstein and colleagues' work (von Kriegstein *et al.*2005).

On the basis of results in Study 2, the updated model predicts that speaker discrimination is processed at least partially in parallel to the voice recognition pathway engaging auditory working memory mechanisms in the inferior parietal cortex. In the original model of voice processing presented in this thesis, accent processing was also hypothesised to recruit inferior parietal and inferior frontal regions in high level explicit judgements, in line with models of emotional prosody (Wildgruber *et al.*2006). Lower level accent discrimination tasks may share cognitive demands with speaker discrimination, and may integrally require working memory to track and compare prosodic contours (either at a segmental or suprasegmental level). It is therefore proposed in the updated model that attentional and working memory mechanisms operate at apperceptive levels as well as at explicit recognition tasks. Neural networks activated by vocal and accent perceptual processes could be compared using fMRI.

7. Clinical implications of voice processing impairments

In everyday life, voice processing deficits are likely to be critical when additional cues are reduced or unavailable such as on the telephone or in the presence of background noise, and particularly in neurodegenerative disease, in conjunction with multimodal impairments of recognition of familiar people and other cognitive impairments, such as episodic memory deficits in AD or executive function deficits in FTLD. Although impairments of voice and accent processing may not be clinically salient symptoms, the work presented in this thesis suggests that this may be at least partly attributable to a lack of assessment of auditory impairments diagnostically and perhaps also to the availability of compensatory cues (from other modalities and context). It is notable that a substantial proportion of carers of patients in both groups in Study 2 had noticed that the patient had some difficulty with voice recognition. Voice processing deficits may contribute importantly to disability, such as social withdrawal in AD (Reichman & Negron 2001) or a loss of social-connectedness characteristic of the behavioural changes associated with right temporal atrophy in FTLD (Chan *et al.*2009; Olson *et al.*2007), behaviours which could be assessed empirically in relation to vocal object and emotion processing.

The data presented in Study 1 demonstrated that phonagnosia can occur with preserved recognition of vocal emotions in a case of bvFTD. However as deficits of affective processing are more commonly described than object recognition impairments in bvFTD it is possible that the reverse dissociation may also occur. Such dissociations may explain heterogeneity in clinical symptoms in this syndrome. Disease-specific deficits (or relative preservation) of voice processing could potentially assist diagnosis; for example differing presentations of vocal processing impairment were demonstrated between cases of bvFTD and tvFTLD in Study 1, between AD and SD groups in Study 2, and between PNFA and AD groups in Study 3. Cognitive deficits in degenerative syndromes may underpin differential patterns of voice impairment; for example in AD deficits of auditory working memory (Stopford, Snowden, Thompson *et al.* 2007) or episodic memory impairment may contribute to deficits of voice processing. Longitudinal study of voice processing performance in relation to other patterns of cognitive impairment is required in patients with degenerative pathologies to understand the evolution of clinical impairments of voice processing.

Appendices

Appendix A.1. Quantification of voice recognition ability: control pilot study.

A pilot study was conducted to quantify voice recognition ability in a sample of older adults firstly to select famous voice stimuli for the main study and secondly to investigate the relationship between voice performance and demographic variables and measures of media exposure. Faces were selected to compare performance in the voice modality to the same identities presented in another non-verbal modality.

A pilot control sample of 26 older controls, 18 female (mean age=65.5, SD=7.3, range: 51-82) were tested using 60 voice and face stimuli, obtained from publicly available sources. Familiarity, naming and recognition of 60 famous and 60 unfamiliar voices and faces were assessed, presenting the same public figures in both modalities. Order of presentation of faces and voices was randomised and balanced between subjects. One control subject from the pilot study was excluded, based on performance greater than 2 standard deviations below the control mean on all tests of voice and face identification; this may reflect a developmental or acquired difficulty with person recognition.

In order to investigate the relationship between background control variables and identification scores, the pilot control sample scores for naming and recognition of the 24 voices and faces used in the final study were combined with scores from the main study control sample. Statistics for the total control sample of 48 controls, 33 female (mean age=64.8, SD=5.9, range: 51-82) were calculated. Voice identification scores (naming: mean=16.2, SD=4.9, range: 5-23; recognition: mean=18.4, SD=4.6, range: 5-24) were lower than scores for identification of the faces (naming: mean=20.1, SD=3.8, range: 8-24; recognition: mean=22.8, SD=2.0, range: 15-24). Paired t-tests were used to assess the difference in control scores between face identification and voice identification scores. Face naming scores were significantly greater than voice naming scores ($t=7.1$, $p<0.001$, $df=47$) and face recognition scores were significantly greater than voice recognition scores ($t=7.9$, $p<0.001$, $df=47$).

In order to assess the relationship between background control variables and voice and face identification test scores, univariate associations between the effects of age, sex, number of years of education, IQ (measured on the NART), naming ability (assessed on the GNT), and media

exposure categorical data: for television watching (hours per week), radio listening (hours per week) and news exposure (see Methods Section 2.2.3) on famous voice naming, voice recognition, face naming and face recognition test scores, respectively. To allow for violations of the normality assumption, bias-corrected bootstrapped confidence intervals (CI) were used.

Associations between voice familiarity, naming and identification tests and background control variables

Demographic variables	**Familiarity** Coefficient (CI)	**Naming** Coefficient (CI)	**Identification** Coefficient (CI)
Age (years)	-0.08 (-0.9, 0.7)	-0.75 (-1.6, 0.1)	-0.55 (-1.4, 0.3)
Gender	1.21 (-9.1, 11.5)	-8.71 (-18.6, 1.2)	-7.15 (-17.3, 1.0)
Years of education	-0.35 (-2.0, 1.3)	0.27 (-1.4, 2.0)	0.36 (-1.5, 2.2.)
NART IQ	0.20 (-0.4, 0.8)	0.47 (-1.0, 1.0)	0.41 (-0.2, 1.0)
GNT (/30)	-0.79 (-2.2., 0.6)	0.77 (-0.4, 2.0)	0.28 (-1.0, 1.6)
Media exposure measures			
Television (hours per week)	3.46 (0.50, 6.4)*	1.81 (-1.4, 5.0)	2.8.6 (-0.18, 5.89)
Radio (hours per week)	1.26 (-2.3., 4.8)	-0.52 (-3.9, 2.9.)	-0.13 (-3.5, 3.3)
News (times per week)	4.00 (2.0, 6.0)**	3.53 (1.2, 5.9)**	3.65 (0.8, 6.5)*

CI represents 95% bootstrapped confidence intervals; Significance level: *p<0.05, **p<0.01

There was no evidence of association between any of the voice recognition scores with background demographic or neuropsychological variables (age, sex, number of years of education, NART IQ, GNT score). There was significant evidence that an increase in exposure to the news was associated with an increase in score on all three voice recognition tests. Hours of television watching were also significantly associated with an increase in voice familiarity score.

To select items for the final study, as the voice was the modality of interest, the 24 public figures best identified from voice by pilot study controls were selected; these items were correctly identified by 64-92% of controls (mean = 75.0, SD=9.0). The same set of public figures presented in the face modality were recognised by a larger proportion (76 – 100%) of controls (mean = 91.8, SD=7.2), and all but one of these 24 public figures (Ann Widdecombe: 88% (voice), 76% (face)) was recognised by a larger proportion of controls in the face modality than the voice modality. Across the entire set of 60 voices, only seven public figures were identified by a larger proportion of controls better from their voice than their faces.

Appendix A.2. List of the background neuropsychological tests used in this Thesis

Background Neuropsychological tests

Test (maximum score)	Test (maximum score)
IQ	**Semantic tests**
WASI Verbal IQ	British Picture Vocabulary Scale (/150)
WASI Performance IQ	Concrete synonyms (/25)
NART full-scale IQ	Abstract synonyms (/25)
Schonell reading IQ	Landmark Naming (/15)
	Landmark Recognition (/15)
Working memory	Famous Faces Test: Naming (/12)
WAIS III Digit span forwards (/14)	
WAIS III Digit span backwards (/14)	**Other non-semantic skills**
WMS-R Digit span forwards (/12)	Graded Naming Test (/30)
WMS-R Digit span backwards (/12)	Object Decision Task (/20)
WMS-III Spatial span forwards (/12)	Graded Difficulty Arithmetic (/24)
WMS-III Spatial span backwards (/12)	
	Executive function
Episodic memory	DKEFS: Stroop Word reading
Recognition Memory Test: words (/50)	DKEFS: Stroop inhibition
Recognition Memory Test: faces (/50)	DKEFS: Design Fluency Task: switching

WASI, Wechsler Abbreviated Scale of Intelligence (Wechsler 1999); NART, National Adult Reading Test (Nelson 1982); Schonell Graded Word Reading Test IQ (Schonell *et al.*1971); BPVS, British Picture Vocabulary Scale (McCarthy *et al.*1992); Concrete and Abstract Synonyms (Warrington, McKenna, & Orpwood 1998); Landmark name, London landmark naming and identification test (Whiteley *et al.*1978); Famous Faces Test: naming and recognition (Warrington *et al.*1967); WAIS III (Wechsler Adult Intelligence Scale- version III) digit span: forwards, backwards (Wechsler 1997); WMS-III (Wechsler Memory Scale-version 3) Spatial Span: forwards, backwards (Wechsler 1999); WMS-R (Wechsler Memory Scale-Revised) digit span: forwards, backwards (Wechsler 1987); Graded Naming Test (Warrington 1997); Object Decision Task (Warrington & James 1991); Graded Difficulty Arithmetic (Jackson & Warrington 1986); Recognition Memory Tests: words and faces (Warrington 1984); DKEFS, Delis-Kaplan Executive Function System Stroop tests and Design Fluency Task (Delis, Kaplan, & Kramer 2001).

Appendix A.3. Lists of the public figures selected for Experiments 1 and 2 and faces frequency matched to voices in Experiment 3

	Experiment 1: main study figures	Experiment 3: frequency matched faces
1	Alan Bennett	Alan Titchmarsh
2	Ann Widdecombe	Anne Robinson
3	Bill Clinton	Anthony Hopkins
4	Billy Connolly	Barbara Windsor
5	Bob Geldof	Bruce Forsyth
6	David Attenborough	Charles Kennedy
7	Edward Heath	Cilla Black
8	George Bush	David Cameron
9	Gordon Brown	David Frost
10	Ian Paisley	David Jason
11	Janet Street-Porter	Dolly Parton
12	Joanna Lumley	Hugh Grant
13	John Humphreys	Jack Nicholson
14	John Major	Jeremy Clarkson
15	Jonathan Ross	Jimmy Carter
16	Judi Dench	John Cleese
17	Kenneth Williams	John Snow
18	Margaret Thatcher	Michael Caine
19	Neil Kinnock	Michael Portillo
20	Prince Charles (written name: Charles Windsor)	Moira Stewart
21	Princess Diana (written name: Diana Spencer)	Robbie Coltrane
22	Ronnie Corbett	Sean Connery
23	Terry Wogan	Stephen Fry
24	Tony Blair	Woody Allen

Appendix A.4.1. Correlations of apperceptive performance: modality and semantic performance

Subtest	tvFTLD n=14	AD n=22
Apperceptive		
Easy speaker discrimination & Benton face test	0.12 (-0.64, 0.66)	0.26 (-0.13, 0.61)
Difficult speaker discrimination & Benton face test	0.16 (-0.37, 0.65)	0.04 (-0.41, 0.59)
Apperceptive & familiarity		
Easy speaker discrimination & voice familiarity	0.07 (-0.46, 0.55)	0.18 (-0.27, 0.67)
Difficult speaker discrimination & voice familiarity	0.42 (-0.15, 0.81)	0.41* (0.05, 0.67)
Benton face test & face familiarity	-0.42 (-0.77, 0.23)	-0.05 (-0.47, 0.35)
Apperceptive & identification		
Easy speaker discrimination & voice identification	0.21 (-0.25, 0.66)	0.10 (-0.33, 0.52)
Difficult speaker discrimination & voice identification	0.23 (-0.73, 0.82)	0.20 (-0.20, 0.56)
Benton face test & face identification	-0.16 (-0.66, 0.35)	0.17 (-0.24, 0.51)

Correlation coefficients are shown with 95% bias-corrected and accelerated bootstrap confidence intervals; *p<0.05

Appendix A.4.2. Correlations between semantic subtests, within modality and between presentation modalities

Subtest	tvFTLD n=14	AD n=22
Within-modality correlations		
Voices		
Familiarity & naming	0.04 (-0.41, 0.66)	0.54* (0.23, 0.72)
Familiarity & identification	0.73 (-0.39, 0.97)	0.67* (0.38, 0.83)
Familiarity & cross-modal	0.70* (0.14, 0.98)	0.81* (0.60, 0.90)
Identification & naming	0.18 (-0.25, 0.88)	0.60* (0.31, 0.78)
Identification & cross-modal	0.84* (0.04, 0.98)	0.78* (0.53, 0.89)
Naming & cross-modal	0.59 (-0.08, 0.99)	0.54* (0.20, 0.74)
Faces		
Familiarity & naming	0.29 (-0.39, 0.73)	0.50* (0.13, 0.75)
Familiarity & identification	0.48 (-0.19, 0.80)	0.67* (0.28, 0.87)
Familiarity & cross-modal	0.53 (-0.15, 0.81)	0.69* (0.30, 0.87)
Identification & naming	0.47* (0.02, 0.94)	0.48* (0.19, 0.70)
Identification & cross-modal	0.89* (0.43, 0.96)	0.78* (0.46, 0.95)
Naming & cross-modal	0.62* (0.22, 0.89)	0.50* (0.18, 0.71)
Between-modality correlations		
Familiarity		
Voice & face	0.08 (-0.56, 0.50)	0.62* (0.30, 0.77)
Voice & name	0.35 (-0.17, 0.76)	0.14 (-0.43, 0.52)
Face & name	-0.01 (-0.52, 0.56)	0.42(-0.20, 0.74)
Naming		
Voice & face	0.94* (0.68, 1.00)	0.62* (0.04, 0.88)
Identification		
Voice & face	0.72* (0.09, 0.87)	0.78* (0.55, 0.89)
Voice & name	0.52 (-0.02, 0.73)	0.51 (-0.02, 0.76)
Face & name	0.82* (0.39, 0.85)	0.75* (0.53, 0.88)
Cross-modal		
Voice & face	0.77* (0.30, 0.96)	0.73* (0.42, 0.90)

Correlation coefficients are shown with 95% bias-corrected and accelerated bootstrap confidence intervals; *p<0.05

Appendix A.4.3. Associations between semantic and perceptual test performance and disease severity measures

	tvFTLD		AD	
	Associations with BPVS	**Associations with disease duration (years)**	**Associations with MMSE**	**Associations with disease duration (years)**
	n=14	**n=14**	**n=22**	**n=22**
Semantic subtest				
Familiarity				
Voice	0.05 (-0.004, 0.10)	-0.34 (-1.38, 0.71)	-0.51 (-1.07, 0.05)	0.09 (-0.75, 0.92)
Face	0.04 (-0.07, 0.08)	0.21 (-2.79, 3.22)	-0.06 (-0.87, 0.75)	0.96 (-0.10, 1.99)
Name	0.08* (0.02, 0.14)	-3.07**(-4.67, -1.47)	0.26 (-0.12, 0.64)	0.34 (-0.13, 0.81)
Naming				
Voice	0.02 (-0.01, 0.04)	-0.35 (-0.88, 0.17)	-0.05 (-0.47, 0.37)	0.52* (0.08,0.97)
Face	0.04 (-0.01, 0.09)	-0.94 (-2.32, 0.44)	0.20 (-0.32, 0.72)	0.99* (0.22, 1.75)
Identification				
Voice	0.05 (-0.01, 0.12)	-0.84 (-1.85, 0.16)	-0.50 (-1.23, 0.22)	0.10 (-0.85, 1.05)
Face	0.11** (0.04, 0.18)	-1.76 (-3.89, 0.37)	-0.31 (-0.86, 0.24)	0.23 (-0.71, 1.17)
Name	0.11** (0.06, 0.16)	-2.46* (-4.59, -0.33)	0.03 (-0.50, 0.56)	0.10 (-0.64, 0.84)
Cross-modal Matching				
Voice	0.10* (0.01, 0.18)	-1.22 (-3.15, 0.71)	-0.51* (-1.00, -0.02)	-0.08 (-1.17, 1.01)
Face	0.10* (0.04, 0.16)	-1.49 (-3.91, 0.92)	-0.34 (-0.73, 0.05)	0.49 (-0.44, 1.41)
Perceptual subtest				
Easy speaker discrimination	0.001 (-0.02, 0.02)	-0.01 (-0.54, 0.52)	0.19 (-0.09, 0.46)	0.30 (-0.14, 0.75)
Difficult speaker discrimination	0.003 (-0.01, 0.02)	-0.12 (-0.39, 0.15)	-0.02 (-0.19, 0.15)	0.30 (-0.05, 0.66)
Benton facial recognition	0.003 (-0.04, 0.04)	0.38 (-0.57, 1.33)	0.56 (-0.1, 1.2)	-0.36 (-1.60, 0.89)

Regression coefficients (95% bootstrapped confidence intervals); Significance level: *p<0.05, **p<0.01

Appendix A.4.4. Correlations between vocal semantic subtests and background neuropsychological performance

	tvFTLD N=14				AD N=22			
Test	**Familiarity**	**Naming**	**Identification**	**Cross modal**	**Familiarity**	**Naming**	**Identification**	**Cross modal**
IQ								
Verbal IQ	**0.40* (0.01, 0.71)**	0.50 (-0.37, 1.0)	0.33 (-0.57, 0.66)	0.51 (-0.12, 0.90)	-0.25 (-0.59, 0.13)	0.12 (-0.39, 0.54)	-0.28 (-0.64, 0.12)	-0.25 (-0.62, 0.05)
Performance IQ	**0.74* (0.40, 0.88)**	-0.16 (-0.60, 0.32)	0.38 (-0.45, 0.86)	0.31 (-0.27, 0.78)	-0.20 (-0.50, 0.25)	0.01 (-0.45, 0.50)	-0.24 (-0.63, 0.30)	-0.08 (-0.47, 0.43)
Reading IQ	0.14 (-0.34, 0.53)	**0.48* (0.03, 0.65)**	0.14 (-0.35, 0.53)	**0.42* (0.002, 0.73)**	-0.22 (-0.54, 0.18)	0.23 (-0.16, 0.52)	-0.38 (-0.69, 0.16)	-0.23 (-0.52, 0.13)
Semantic tests								
BPVS	0.51 (-0.05, 0.85)	**0.47* (0.11, 0.87)**	0.67 (-0.31, 0.76)	**0.67* (0.19. 0.90)**	0.20 (-0.13, 0.53)	**0.34* (0.03, 0.55)**	0.001 (-0.51, 0.49)	0.19 (-0.20, 0.70)
Concrete synoynms[c]	0.004 (-0.78, 0.60)	**0.63* (0.00, 0.98)**	0.21 (-0.92, 0.79)	0.49 (-0.28, 0.93)	0.06 (-0.34, 0.46)	0.25 (-0.18, 0.56)	-0.02 (-0.35, 0.42)	-0.08 (-0.48, 0.32)
Abstract synonyms[c]	0.14 (-0.79, 0.71)	0.45 (-0.29, 1.0)	0.37 (-0.67, 0.75)	0.47 (-0.31, 0.92)	-0.10 (-0.39, 0.27)	0.27 (-0.11, 0.55)	-0.09 (-0.51, 0.27)	-0.07 (-0.42, 0.29)
Landmark name[d]	-0.10 (-0.65, 0.55)	0.85 (-0.03, 1.0)	-0.003 (-0.57, 0.85)	0.36 (-0.40, 0.97)	0.14 (-0.40, 0.58)	**0.58* (0.15, 0.78)**	0.22 (-0.13, 0.57)	0.14 (-0.38, 0.60)
Landmark recogn[d]	0.05 (-0.58, 0.74)	0.54 (-0.22, 0.97)	-0.01 (-0.65, 0.72)	0.16 (-0.40, 0.85)	0.33 (-0.22, 0.67)	**0.67* (0.45, 0.81)**	**0.53* (0.13, 0.76)**	0.39 (-0.09, 0.74)
Other non-semantic skills								
GNT	0.13 (-0.70, 0.46)	**0.96* (0.14, 1.0)**	0.24 (-0.43, 0.79)	**0.62* (0.05, 0.98)**	0.09 (-0.27, 0.44)	**0.44* (0.06, 0.70)**	0.01 (-0.38, 0.39)	0.05 (-0.37, 0.48)
Object Decision task	0.34 (-0.48, 0.71)	0.11 (-0.40, 0.43)	0.15 (-0.31, 0.61)	0.27 (-0.05, 0.57)	0.05 (-0.39, 0.43)	-0.07 (-0.45, 0.32)	-0.19 (-0.57, 0.29)	-0.13 (-0.51, 0.37)
Digit span fwd	-0.12 (-0.72, 0.51)	**0.47* (0.0, 0.70)**	-0.15 (-0.68, 0.69)	0.06 (-0.57, 0.71)	-0.16 (-0.53, 0.17)	-0.04 (-0.48, 0.40)	-0.18 (-0.51,0.18)	-0.31 (-0.59, 0.03)
Digit span back	0.29 (-0.47, 0.68)	0.04 (-0.69, 0.26)	0.05 (-0.38, 0.46)	0.10 (-0.42, 0.54)	-0.22 (-0.54, 0.12)	0.002 (-0.54, 0.59)	-0.26 (-0.64, 0.23)	-0.37 (-0.70, 0.05)
Arithmetic (GDA)	0.38 (-0.16, 0.84)	0.24 (-0.09, 0.50)	0.07 (-0.31, 0.42)	0.24 (-0.20, 0.60)	-0.15 (-0.52, 0.28)	0.03 (-0.43, 0.43)	-0.37 (-0.69, 0.18)	-0.26 (-0.62, 0.21)
Episodic memory								
RMT words[b]	0.28 (-0.19, 0.58)	0.32 (-0.21, 0.58)	0.25 (-0.20, 0.54)	0.36 (-0.12, 0.67)	0.09 (-0.33, 0.48)	0.21 (-0.26, 0.58)	**0.45** (0.12, 0.70)**	0.16 (-0.30, 0.49)
RMT faces[b]	0.59 (-0.39, 0.79)	0.07 (-0.30, 0.43)	0.20 (-0.41, 0.75)	0.25 (-0.23, 0.75)	**0.48* (0.06, 0.78)**	0.16 (-0.34, 0.61)	0.32 (-0.04, 0.64)	0.33 (-0.10, 0.66)
Executive function								
Stroop Inhibition scaled[a]	0.05 (-0.71, 0.69)	-0.16 (-0.53, 0.14)	-0.33 (-0.72, 0.27)	-0.31 (-0.73, 0.32)	-0.09 (-0.43, 0.27)	0.05 (-0.37, 0.62)	-0.38 (-0.69, 0.13)	-0.32 (-0.65, 0.18)

Correlation coefficients are shown with 95% bias-corrected and accelerated bootstrap confidence intervals; Significant correlations are shown in bold (*p<0.05).WASI verbal and performance IQ; WMS-R digit span forwards and backwards; GDA, Graded Difficulty Arithmetic; Landmark name, London landmark naming and recognition tests; RMT, Recognition Memory Tests; DKEFS Stroop test switched scaled score; WASI, Wechsler Abbreviated Scale of Intelligence; [a]n=12 (2 tvFTLD subjects were unable to name colours); [b]1 tvFTLD subject did not perform recognition memory tasks. [c] 2 tvFTLD and 1 AD subject did not perform synonyms tests. [d]3 tvFTLD and 2AD subjects did not perform the London landmarks test.

Appendix A.4.5. Correlations between speaker discrimination and neuropsychological performance

Test	tvFTLD N=14		AD N=22	
	Easy speaker	**Difficult speaker**	**Easy speaker**	**Difficult speaker**
Performance IQ	0.18 (-0.30, 0.61)	**0.62* (0.29, 0.83)**	0.30 (-0.10, 0.61)	0.04 (-0.35, 0.50)
Digit span fwd	-0.01 (-0.42, 0.46)	-0.28 (-0.74, 0.28)	**0.34*(0.01, 0.66)**	0.24 (-0.15, 0.58)
Digit span back	-0.14 (-0.65, 0.36)	0.27 (-0.26, 0.66)	0.25 (-0.24, 0.64)	0.20 (-0.16, 0.59)
BPVS	0.04 (-0.46, 0.46)	0.11 (-0.64, 0.67)	-0.03 (-0.28, 0.26)	-0.26 (-0.47, 0.09)
GNT	-0.10 (-0.60, 0.31)	-0.50 (-0.86, 0.30)	0.09 (-0.24, 0.40)	0.10 (-0.37, 0.50)
RMT words[b]	0.32 (-0.22, 0.67)	0.15 (-0.47, 0.60)	-0.06 (-0.55, 0.39)	0.12 (-0.36, 0.58)
RMT faces[b]	-0.01 (-0.60, 0.64)	0.09 (-0.33, 0.69)	0.20 (-0.32, 0.71)	0.24 (-0.20, 0.59)
Arithmetic (GDA)	-0.25 (-0.66, 0.26)	0.004 (-0.43, 0.44)	0.23 (-0.28, 0.58)	0.18 (-0.22, 0.54)

Correlation coefficients are shown with 95% bias-corrected and accelerated bootstrap confidence intervals; Significant correlations are shown in bold (*p<0.05). WASI Performance IQ; WMS-R digit span; GDA, Graded Difficulty Arithmetic; RMT, Recognition Memory Tests; WASI, Wechsler Abbreviated Scale of Intelligence; [a]n=12 (2 tvFTLD subjects were unable to name colours); [b]1 tvFTLD subject did not perform recognition memory tasks.

Appendix A.4.6. Number of patients (and proportion of each patient group) impaired at 0, 1, 2 & 3 modalities of presentation on familiarity, identification, naming and cross-modal recognition semantic tasks

	Familiarity		**Identification**		**Naming**		**Cross-modal recognition**	
	FTLD n=14	AD n=22	FTLD n=14	ADs n=22	FTLD n=14	AD n=22	FTLD n=14	AD n=22
Number of modalities impaired	No. of patients % of patient group		No. of patients % of patient group		No. of patients % of patient group		No. of patients % of patient group	
0 modalities	0 0.00%	9 40.9%	0 0.00%	5 22.7%	0 0%	1 4.5%	0 0%	5 22.7%
1 modality								
Voice only	1 7.1%	0 0.00%	0 0.00%	1 4.6%	0 0%	1 4.5%	0 0%	0 0%
Face only	1 7.1%	5 22.7%	0 0.00%	0 0.00%	0 0%	0 0%	1 7.1%	3 13.6%
Name only	0 0.00%	0 0.00%	1 7.1%	1 4.6%	-	-	-	-
2 modalities								
Voice & face	1 7.1%	6 27.3%	0 0.00%	2 9.1%	14 100%	20 90.9%	13 92.9%	14 63.6%
Voice & name	1 7.1%	0 0.00%	0 0.00%	1 4.6%	- -	- -	- -	- -
Face & name	0 0.00%	0 0.00%	0 0.00%	2 9.1%	- -	- -	- -	- -
All 3 modalities	10 71.4%	2 9.1%	13 92.9%	10 45.5%	-	-	-	-

- Cross modal and naming tasks were not performed in the name modality

Appendix A.4.7. Comparison of right-sided versus left-sided tvFTLD subgroups

	Right-sided tvFTLD N=4	**Left-sided tvFTLD N=9**	**Right - left sided:**	**Right – Left Direction of difference in means**
	Mean (SD)	Mean (SD)	Difference in means (95% CI)	
Demographics				
Age (years)	63.0 (6.2)	63.9 (6.4)	-0.9 (-8.0, 6.3)	-
Years of education	13.0 (3.8)	13.0 (3.9)	0.0 (-2.9, 10.4)	
Disease duration (years)	5.0 (2.2)	5.3 (1.3)	0.4 (-2.2, 3.0)	+
General neuropsychological Test (max score)				
MMSE score (/30)	25.3 (3.8)	18.9 (8.0)	**6.1* (0.1, 12.1)**	+
Verbal IQ	75.8 (25.7)	59.2 (13.1)	**30.5* (3.6, 57.5)**	+
Performance IQ	98.5 (7.1)	98.6 (23.1)	3.7 (-14.9, 22.3)	+
Reading IQ	94.8 (19.6)	82.5 (24.3)	20.8 (-5.9, 47.4)	+
GNT (/30)	5.8 (11.5)	0.2 (0.7)	6.3 (-4.0, 16.7)	+
BPVS (/150)	84.3 (47.9)	61.1 (47.5)	45.6 (-12.5, 103.8)	+
Object decision task (/20)	13.5 (3.9)	17.7 (5.4)	-1.7 (-5.4, 2.1)	-
RMT words (/50)	37.0 (7.1)	34.5 (7.7)	2.8 (-5.3, 10.8)	+
RMT faces (/50)	27.5 (2.6)	29.6 (4.9)	-0.9 (-4.3, 2.5)	-
Experimental perceptual tasks				
Easy speaker discrimination (/28)	24.5 (1.9)	24.7 (1.7)	0.8 (-0.6, 2.2)	+
Hard speaker discrimination (/12)	9.0 (1.4)	9.2 (1.2)	0.0 (-1.6, 1.6)	
Gender (/24)	24.0 (0.0)	24.0 (0.0)	0.0 (-, -) [a]	
Size (/20)	17.8 (3.2)	16.0 (2.6)	1.8 (-1.7, 5.2)	+
Benton (/56)	47.3 (2.2)	41.0 (3.2)	**4.0* (0.6, 7.4)**	+
Experimental semantic tasks				
Voice				
Voice familiarity (/48)	27.7 (6.0)	27.0 (1.8)	-0.7 (-4.8, 3.4)	-
Voice naming (/24)	1.5 (3.0)	0.2 (0.4)	1.3 (-1.6, 4.2)	+
Voice identification (/24)	2.5 (3.0)	3.0 (6.1)	-0.5 (-5.2, 4.2)	-
Voice cross-modal (/24)	6.5 (9.7)	6.7 (6.8)	0.3 (-9.7, 10.3)	+
Face				
Face familiarity (/48)	30.3 (9.9)	36.9 (5.2)	-5.9 (-15.9, 4.1)	-
Face naming (/24)	4.0 (6.7)	1.3 (2.1)	2.9 (-3.5, 9.3)	+
Face identification (/24)	7.5 (7.9)	8.2 (8.1)	-0.7 (-9.9, 8.4)	-
Cross-modal matching (/24)	8.3 (9.3)	10.9 (7.7)	-1.4 (-11.4, 8.7)	-
Name				
Name familiarity (/48)	35.5 (6.8)	34.4 (8.1)	0.3 (-8.0, 8.6)	+
Name identification (/24)	7.3 (8.8)	7.1 (7.5)	1.4 (-7.9, 10.7)	+

CI represents 95% bootstrapped confidence intervals; *Significant results are shown in bold, and indicate where left-sided group significantly worse than right-sided group at $p<0.05$ [a]All subjects performed at ceiling on this test, therefore bootstrapped CI could not be calculated. WASI verbal and performance IQ; RMT, Recognition Memory Tests.

Appendix A.5.1. Spoken sentences in question comprehension test

	Questions	Examples of accepted answers
1	What is the opposite of young?	Old
2	What is rain made of?	Water
3	What hand do you write with?	right /left
4	What is the hottest time of the year?	summer/July
5	What colour is butter?	yellow/golden
6	What shape is the earth?	round/spherical
7	How many inches in a foot?	Twelve
8	What is 3 times 10?	Thirty
9	Can a bird fly?	Yes
10	What does a bell do?	rings/clangs
11	What is the opposite of good?	bad/evil
12	What do you sleep in?	bed/pyjamas
13	What room do you cook in?	kitchen
14	From what animal do we get milk?	cow
15	What does a key open?	door/lock
16	Where do you wear a ring?	finger
17	What are windows made of?	glass
18	What is the opposite of white?	black
19	What colour is the sky?	blue/grey
20	What do girls grow up to be?	women/ladies
21	What is the opposite of long?	short
22	What colour is blood?	red
23	Where does a picture hang?	wall/gallery
24	What is a rose?	flower
25	What's another word for cash?	money/dosh
26	What does a chicken lay?	egg(s)
27	What do we hear with?	ears
28	What are nails made of?	metal/keratin
29	Who goes to school?	children/pupils
30	What does an honest man always tell?	truth
31	What does a watch tell you?	time
32	What do you burn on an open fire?	wood/coal
33	How many legs does a dog have?	4
34	What do you find in a library?	books/computers
35	What does a bird build?	nest
36	Who do you see when you're ill?	doctor/nurse
37	How many pennies in the pound?	100/240
38	What city are we in?	London
39	What is a very young child called?	baby/ infant
40	When can you see the moon?	night/evening

Appendix A.5.2. Stimuli used in the word verification task

	Target	Distractor	Change
1	dame	name	1st Consonant
2	neat	meat	1st Consonant
3	night	might	1st Consonant
4	nip	lip	1st Consonant
5	pail	tail	1st Consonant
6	pill	bill	1st Consonant
7	tack	sack	1st Consonant
8	bag	back	Last consonant
9	bean	beam	Last consonant
10	cab	cap	Last consonant
11	code	coat	Last consonant
12	maid	main	Last consonant
13	bad	bed	Vowel
14	bat	bit	Vowel
15	cat	cut	Vowel
16	deed	dead	Vowel
17	fall	full	Vowel
18	gut	get	Vowel
19	mat	met	Vowel
20	pit	pet	Vowel
21	rice	race	Vowel
22	slap	slip	Vowel
23	tap	tip	Vowel
24	tape	type	Vowel

Word pairs derived from the PALPA Minimal Pairs test

Appendix A.5.3. Stimulus trials in the regional accent recognition tests

	Regional British accents			Regional English accents	
Trial	**Answer**	**Region**		**Answer**	**Region**
1	Wales	Glamorgan	1	North	Lancashire
2	England	Merseyside	2	North	Merseyside
3	Ireland	Londonderry	3	South	Essex
4	Scotland	Scottish borders	4	South	Oxfordshire
5	England	Essex	5	South	Hackney
6	England	Yorkshire	6	South	Devon
7	Scotland	Edinburgh	7	North	Lancashire
8	Wales	Port Talbot	8	South	Wiltshire
9	England	Standard English	9	South	Norfolk
10	Ireland	Limerick	10	North	Northumberland
11	Ireland	Belfast	11	South	Gloucestershire
12	Wales	Glamorgan	12	North	Merseyside
13	Scotland	Glasgow	13	North	Yorkshire
14	Wales	Swansea	14	North	Merseyside
15	Scotland	North Ayrshire	15	South	Kent
16	Ireland	Cork	16	North	Yorkshire
17	Wales	Pembrokeshire	17	North	Yorkshire
18	Ireland	County Antrim	18	South	Bristol
19	Scotland	Glasgow	19	North	Durham
20	England	Standard English	20	South	East Sussex
21	England	Yorkshire	21	South	Oxfordshire
22	Scotland	Aberdeenshire	22	North	Cheshire
23	Wales	Powys	23	North	Tyne and Wear
24	Ireland	Armagh	24	South	East Sussex

Appendix A.5.4. Stimuli used in the test of naming of countries from verbal description

	Country	Stimulus
1	Wales	What country does a leek represent?
2	Ireland	Which country does Guinness come from?
3	India	Which country did Ghandi come from?
4	Germany	Which country did Hitler lead?
5	Spain	Which country is famous for flamenco and bullfighting?
6	France	Of what country was De Gaulle the president?
7	England	In which country are London and Birmingham?
8	Scotland	From which country does the haggis come?
9	America/USA	The stars and stripes is the flag of which country?
10	Italy	Spaghetti comes from which country?

Appendix A.5.5. Correlations between accent comprehension tests and phoneme discrimination, and between tests of accent recognition and country knowledge tests within the AD group (N=20)

	Accent comprehension		Accent recognition		
	Questions Foreign – English difference score	Verification Foreign – English difference score	British regions	English regions	England versus Foreign (Block 1)
Phoneme discrimination					
Minimal pair word verification	0.22 (-0.18, 0.77)	-0.21 (-0.74, 0.13)	-	-	-
Country knowledge tests					
Country naming from description	-	-	**0.52 (0.09, 0.84)***	0.18 (-0.26, 0.52)	0.22 (-0.18, 0.68)
Map naming	-	-	0.36 (-0.27, 0.72)	-0.13 (-0.52 0.27)	0.0 (-0.50, 0.42)
Map comprehension	-	-	0.44 (-0.11, 0.77)	-0.20 (-0.59, 0.26)	0.03 (-0.38, 0.44)

Correlation coefficients are shown with 95% bias-corrected and accelerated bootstrap confidence intervals; *$p<0.05$

Appendix A.5.6. Correlations between accent processing tests and subset of background neuropsychological tests, and apperceptive and semantic voice processing tests within the AD group (N=20)

	Accent comprehension		Accent recognition		
	Questions Foreign – English difference score	Verification Foreign – English difference score	British regions	English regions	England versus Foreign (Block 1)
Neuropsychological tests					
Performance IQ	0.34 (-0.12, 0.70)	-0.05 (-0.49, 0.32)	**0.42 (0.01, 0.72)***	-0.32 (-0.67, 0.15)	0.15 (-0.22, 0.50)
BPVS	**0.70 (0.19, 0.93)***	0.19 (-0.20, 0.45)	**0.39 (0.03, 0.80)***	0.07 (-0.23, 0.39)	0.09 (-0.23, 0.44)
Digit span fwd	**0.55 (0.14, 0.81)***	-0.02 (-0.53, 0.45)	0.21 (-0.32, 0.47)	-0.14 (-0.54, 0.33)	0.06 (-0.34, 0.38)
Digit span back	0.15 (-0.28, 0.50)	-0.06 (-0.64, 0.37)	0.08 (-0.50, 0.62)	-0.47 (-0.79, 0.03)	0.19 (-0.21, 0.44)
RMT words	-0.09 (-0.51, 0.47**)**	-0.02 (-0.54, 0.55)	**0.44 (0.14, 0.63)***	0.22 (-0.36, 0.69)	0.29 (-0.13, 0.59)
RMT faces	0.01 (-0.29, 0.43)	-0.17 (-0.69, 0.46)	0.43 (-0.09, 0.75)	0.30 (-0.31, 0.69)	**0.64 (0.10, 0.85)***
Stroop Switching Scaled	0.34 (-0.08, 0.56)	-0.07 (-0.48, 0.40)	0.13 (-0.27, 0.58)	-0.16 (-0.61, 0.32)	0.39 (-0.02, 0.65)
Voice perception					
Easy speaker discrimination	0.06 (-0.27, 0.45)	**-0.72 (-0.89, -0.37)***	0.36 (-0.13, 0.75)	-0.17 (-0.75, 0.57)	0.35 (-0.20, 0.83)
Hard speaker discrimination	-0.09 (-0.45, 0.41)	-0.16 (-0.66, 0.40)	0.004 (-0.37, 0.35)	0.16 (-0.43, 0.61)	**0.50 (0.167, 0.77)***
Familiar voice recognition					
Familiarity	-0.16 (-0.46, 0.20)	-0.17 (-0.68, 0.28)	**0.53 (0.24, 0.77)***	**0.61 (0.32, 0.82)***	**0.72 (0.37, 0.89)***
Naming	-0.38 (-0.68, 0.05)	0.36 (-0.18, 0.72)	**0.58 (0.17, 0.84)***	**0.61 (0.30, 0.82)***	**0.52 (0.24, 0.72)***
Identification	0.26 (-0.23, 0.65)	-0.13 (-0.48, 0.25)	**0.62 (0.31, 0.81)***	**0.52 (0.16, 0.76)***	0.34 (-0.07, 0.65)
Cross-modal matching	0.04 (-0.46, 0.40)	-0.15 (-0.59, 0.29)	**0.66 (0.36, 0.86)***	**0.56 (0.23, 0.78)***	0.44 (-0.03, 0.77)

Correlation coefficients are shown with 95% bias-corrected and accelerated bootstrap confidence intervals; * Significant correlations are shown in bold (*p<0.05); WASI Performance IQ; WMS-R digit span: forwards, backwards; GDA, Graded Difficulty Arithmetic; RMT, Recognition Memory Tests; DKEFs Stroop switching scaled.

Appendix A.6. Publications arising from this Thesis

1. Hailstone JC, Crutch SJ, Vestergaard MD, Patterson RD, Warren JD. Progressive associative phonagnosia: a neuropsychological analysis. *Neuropsychologia,* 48: 1104-1114, 2010.

2. Hailstone JC, Crutch SJ, Warren JD. Voice recognition in dementia. *Behavioural Neurology,* 23: 163-164, 2010.

3. Hailstone JC, Ridgway GR, Bartlett JW, Goll JC, Buckley AH, Crutch SJ, Warren JD. Voice processing in dementia: a neuropsychological and neuroanatomical analysis. *Brain,* 134: 2535-2547, 2011.

4. Hailstone JC, Ridgway GR, Bartlett JW, Goll JC, Crutch SJ, Warren JD. Accent processing in dementia. *Neuropsychologia* 50: 2233-2244, 2012.

Reference List

Acer N, Sahin B, Bas O, Ertekin T, Usanmaz M. Comparison of three methods for the estimation of total intracranial volume: stereologic, planimetric, and anthropometric approaches. *Annals of Plastic Surgery,* 58: 48-53, 2007.

Acres K, Taylor KI, Moss HE, Stamatakis EA, Tyler LK. Complementary hemispheric asymmetries in object naming and recognition: a voxel-based correlational study. *Neuropsychologia,* 47: 1836-1843, 2009.

Adank P, Evans BG, Stuart-Smith J, Scott SK. Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology: Human Perception and Performance,* 35: 520-529, 2009.

Adank P, Janse E. Comprehension of a novel accent by young and older listeners. *Psychology and Aging,* 25: 736-740, 2010.

Adank P, Noordzij ML, Hagoort P. The role of planum temporale in processing accent variation in spoken language comprehension. *Human Brain Mapping,* 33: 360-372, 2012.

Adolphs R, Damasio H, Tranel D, Cooper G, Damasio AR. A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *Journal of Neuroscience,* 20: 2683-2690, 2000.

Allender J, Kaszniak AW. Processing of emotional cues in patients with dementia of the Alzheimer's type. *International Journal of Neuroscience,* 46: 147-155, 1989.

Andics A, McQueen JM, Petersson KM, Gal V, Rudas G, Vidnyanszky Z. Neural mechanisms for voice recognition. *Neuroimage,* 52: 1528-1540, 2010.

Ash S, McMillan C, Gunawardena D, Avants B, Morgan B, Khan A, Moore P *et al*. Speech errors in progressive non-fluent aphasia. *Brain and Language,* 113: 13-20, 2010.

Ashburner J, Friston KJ. Voxel-based morphometry--the methods. *Neuroimage,* 11: 805-821, 2000.

Ashburner J. A fast diffeomorphic image registration algorithm. *Neuroimage,* 38: 95-113, 2007.

Ashburner J, Friston KJ. Computing average shaped tissue probability templates. *Neuroimage,* 45: 333-341, 2009.

Assal G, Aubert C, Buttet J. Cerebral Asymmetry and Voice Recognition. *Revue Neurologique,* 137: 255-268, 1981.

Baayen RH, Piepenbrock R, Gulikers L. The CELEX Lexical Database (CD-ROM). Pennsylvania: Linguistic Data Consortium, University of Pennsylvania; 1995.

Baird A, Samson S. Memory for music in Alzheimer's disease: unforgettable? *Neuropsychology Review,* 19: 85-101, 2009.

Balthazar ML, Yasuda CL, Pereira FR, Bergo FP, Cendes F, Damasceno BP. Coordinated and circumlocutory semantic naming errors are related to anterolateral temporal lobes in mild AD, amnestic mild cognitive impairment, and normal aging. *Journal of the International Neuropsychological Society,* 16: 1099-1107, 2010.

Barnes J, Ourselin S, Fox NC. Clinical application of measurement of hippocampal atrophy in degenerative dementias. *Hippocampus,* 19: 510-516, 2009.

Barrett AM, Crucian GP, Raymer AM, Heilman KM. Spared comprehension of emotional prosody in a patient with global aphasia. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology,* 12: 117-120, 1999.

Baumann O, Belin P. Perceptual scaling of voice identity: common dimensions for different vowels and speakers. *Psychological Research,* 74: 110-120, 2010.

Bayard D, Gallois C, Weatherall A, Pittam J. Pax Americana? Accent attitudinal evaluations in New Zealand, Australia and America. *Journal of Sociolinguistics,* 5: 22-49, 2001.

Beauchemin M, De Beaumont L, Vannasing P, Turcotte A, Arcand C, Belin P, Lassonde M. Electrophysiological markers of voice familiarity. *European Journal of Neuroscience,* 23: 3081-3086, 2006.

Beauchemin M, Gonzalez-Frankenberger B, Tremblay J, Vannasing P, Martinez-Montes E, Belin P, Beland R *et al*. Mother and stranger: an electrophysiological study of voice processing in newborns. *Cerebral Cortex,* 21: 1705-1711, 2011.

Bedard C, Belin P. A "voice inversion effect". *Brain and Cognition,* 55: 247-249, 2004.

Belin P, Zatorre RJ. 'What', 'where' and 'how' in auditory cortex. *Nature Neuroscience,* 3: 965-966, 2000.

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. Voice-selective areas in human auditory cortex. *Nature,* 403: 309-312, 2000.

Belin P, Zatorre RJ, Ahad P. Human temporal-lobe response to vocal sounds. *Brain Research Cognitive Brain Research,* 13: 17-26, 2002.

Belin P, Zatorre RJ. Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport,* 14: 2105-2109, 2003.

Belin P, Fecteau S, Bedard C. Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences,* 8: 129-135, 2004.

Belizaire G, Fillion-Bilodeau S, Chartrand JP, Bertrand-Gauvin C, Belin P. Cerebral response to 'voiceness': a functional magnetic resonance imaging study. *Neuroreport,* 18: 29-33, 2007.

Benton AL, Hamsher KS, Varney N, Spreen O. Contributions to neuropsychological assessment: a clinical manual. Oxford: Oxford University Press; 1989.

Benzagmout M, Chaoui MF, Duffau H. Reversible deficit affecting the perception of tone of a human voice after tumour resection from the right auditory cortex. *Acta Neurochirurgica,* 150: 589-593, 2008.

Berman SM, Mandelkern MA, Phan H, Zaidel E. Complementary hemispheric specialization for word and accent detection. *Neuroimage,* 19: 319-331, 2003.

Best CT, McRoberts GW, Sithole NM. Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance,* 14: 345-360, 1988.

Best CT, McRoberts GW, Goodell E. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America,* 109: 775-794, 2001.

Bilecen D, Seifritz E, Scheffler K, Henning J, Schulte AC. Amplitopicity of the human auditory cortex: an fMRI study. *Neuroimage,* 17: 710-718, 2002.

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET. Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex,* 10: 512-528, 2000.

Binney RJ, Embleton KV, Jefferies E, Parker GJ, Ralph MA. The ventral and inferolateral aspects of the anterior temporal lobe are crucial in semantic memory: evidence from a novel direct comparison of distortion-corrected fMRI, rTMS, and semantic dementia. *Cerebral Cortex,* 20: 2728-2738, 2010.

Bishop CW, Miller LM. A multisensory cortical network for understanding speech in noise. *Journal of Cognitive Neuroscience,* 21: 1790-1805, 2009.

Blumstein SE, Alexander MP, Ryalls JH, Katz W, Dworetzky B. On the nature of the foreign accent syndrome: a case study. *Brain and Language,* 31: 215-244, 1987.

Boeve BF, Geda YE. Polka music and semantic dementia. *Neurology,* 57: 1485, 2001.

Bozeat S, Lambon Ralph MA, Patterson K, Garrard P, Hodges JR. Non-verbal semantic impairment in semantic dementia. *Neuropsychologia,* 38: 1207-1215, 2000.

Bradlow AR, Bent T. Perceptual adaptation to non-native speech. *Cognition,* 106: 707-729, 2008.

Brambati SM, Rankin KP, Narvid J, Seeley WW, Dean D, Rosen HJ, Miller BL *et al*. Atrophy progression in semantic dementia with asymmetric temporal involvement: a tensor-based morphometry study. *Neurobiology of Aging,* 30: 103-111, 2009.

Bright P, Moss H, Tyler LK. Unitary vs multiple semantics: PET studies of word and picture processing. *Brain and Language,* 89: 417-432, 2004.

Bruce V, Valentine T. Identity priming in the recognition of familiar faces. *British Journal of Psychology,* 76 (Pt 3): 373-383, 1985.

Bruce V, Young A. Understanding face recognition. *British Journal of Psychology,* 77: 305-327, 1986.

Buchanan TW, Lutz K, Mirzazade S, Specht K, Shah NJ, Zilles K, Jancke L. Recognition of emotional prosody and verbal components of spoken language: an fMRI study. *Brain Research Cognitive Brain Research,* 9: 227-238, 2000.

Buchsbaum BR, Hickok G, Humphries C. Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cognitive Science,* 25: 663-678, 2001.

Buckner RL, Sepulcre J, Talukdar T, Krienen FM, Liu HS, Hedden T, Andrews-Hanna JR *et al*. Cortical Hubs Revealed by Intrinsic Functional Connectivity: Mapping, Assessment of Stability, and Relation to Alzheimer's Disease. *Journal of Neuroscience,* 29: 1860-1873, 2009.

Burda AN, Scherz JA, Hageman CF, Edwards HT. Age and understanding speakers with Spanish or Taiwanese accents. *Perceptual and Motor Skills,* 97: 11-20, 2003.

Burda AN, Hageman CF, Brousard KT, Miller AL. Dementia and identification of words and sentences produced by native and nonnative English speakers. *Perceptual and Motor Skills,* 98: 1359-1362, 2004.

Burda AN, Brace A, Hosch A. Aphasia and accent in identifying medical sentences. *Perceptual and Motor Skills,* 104: 1375-1378, 2007.

Burda AN, Bradley-Potter M, Dralle J, Murphy J, Ries S, Roehs A. Influence of Age and Native Language on Immediate Verbal Repetition. *Perceptual and Motor Skills,* 109: 169-176, 2009.

Burton AM, Bruce V, Johnston RA. Understanding face recognition with an interactive activation model. *British Journal of Psychology,* 81: 361-380, 1990.

Burton AM, Bruce V. Naming faces and naming names: exploring an interactive activation model of person recognition. *Memory,* 1: 457-480, 1993.

Burton MW, Small SL, Blumstein SE. The role of segmentation in phonological processing: An fMRI investigation. *Journal of Cognitive Neuroscience,* 12: 679-690, 2000.

Busigny T, Robaye L, Dricot L, Rossion B. Right anterior temporal lobe atrophy and person-based semantic defect: a detailed case study. *Neurocase,* 15: 485-508, 2009.

Campanella S, Belin P. Integrating face and voice in person perception. *Trends in Cognitive Sciences,* 11: 535-543, 2007.

Chan D, Anderson V, Pijnenburg Y, Whitwell J, Barnes J, Scahill R, Stevens JM *et al*. The clinical profile of right temporal lobe atrophy. *Brain,* 132: 1287-1298, 2009.

Chang EF, Rieger JW, Johnson K, Berger MS, Barbaro NM, Knight RT. Categorical speech representation in human superior temporal gyrus. *Nature Neuroscience,* 13: 1428-U169, 2010.

Chartrand JP, Belin P. Superior voice timbre processing in musicians. *Neuroscience Letters,* 405: 164-167, 2006.

Chartrand JP, Peretz I, Belin P. Auditory recognition expertise and domain specificity. *Brain Research,* 1220: 191-198, 2008.

Childers DG, Wu K. Gender recognition from speech. Part II: Fine analysis. *Journal of the Acoustical Society of America,* 90: 1841-1856, 1991.

Clarke CM, Garrett MF. Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America,* 116: 3647-3658, 2004.

Clopper CG, Pisoni DB. Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics,* 32: 111-140, 2004a.

Clopper CG, Pisoni DB. Effects of talker variability on perceptual learning of dialects. *Language and Speech,* 47: 207-239, 2004b.

Clopper CG, Pisoni DB. Free classification of regional dialects of American English. *Journal of Phonetics,* 35: 421-438, 2007.

Clopper CG, Bradlow AR. Perception of dialect variation in noise: intelligibility and classification. *Language and Speech,* 51: 175-198, 2008.

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. *Cognitive Neuropsychology,* 21: 513-527, 2004.

Compton AJ. Effects of filtering and vocal duration upon identification of speakers, aurally. *Journal of the Acoustical Society of America,* 35: 1748-1752, 1963.

Crawford JR, Howell DC. Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist,* 12: 482-486, 1998.

Crawford JR, Garthwaite PH. Testing for suspected impairments and dissociations in single-case studies in neuropsychology: evaluation of alternatives using monte carlo simulations and revised tests for dissociations. *Neuropsychology,* 19: 318-331, 2005.

Croot K, Patterson K, Hodges JR. Single word production in nonfluent progressive aphasia. *Brain and Language,* 61: 226-273, 1998.

Crutch SJ, Warrington EK. Spatial coding of semantic information: knowledge of country and city names depends on their geographical proximity. *Brain,* 126: 1821-1829, 2003.

Crutch SJ, Warrington EK. The semantic organisation of proper nouns: the case of people and brand names. *Neuropsychologia,* 42: 584-596, 2004.

Crutch SJ, Warrington EK. Spatially coded semantic information about geographical terms. *Neuropsychologia,* 48: 2120-2129, 2010.

Cummings KE, Chin SB, Pisoni DB. Analysis of the glottal excitation of intoxicated versus sober speech: a first report. *Journal of the Acoustical Society of America,* 99: 2549-2574, 1996.

Damasio AR. Category-related recognition defects as a clue to the neural substrates of knowledge. *Trends in Neurosciences,* 13: 95-98, 1990.

Damjanovic L, Hanley JR. Recalling episodic and semantic information about famous faces and voices. *Memory and Cognition,* 35: 1205-1210, 2007.

de Mareuil PB, Vieru-Dimulescu B. The contribution of prosody to the perception of foreign accent. *Phonetica,* 63: 247-267, 2006.

De Renzi E, Faglioni P, Grossi D, Nichelli P. Apperceptive and associative forms of prosopagnosia. *Cortex,* 27: 213-221, 1991.

De RE, Bonacini MG, Faglioni P. Right posterior brain-damaged patients are poor at assessing the age of a face. *Neuropsychologia,* 27: 839-848, 1989.

Delis DC, Kaplan E, Kramer JH. Delis-Kaplan Executive Function System. San Antonio: The Psychological Corporation; 2001.

della Rocchetta AI, Cipolotti L, Warrington EK. Countries: Their selective impairment and selective preservation. *Neurocase,* 4: 99-109, 1998.

Demonet JF, Chollet F, Ramsay S, Cardebat D, Nespoulous JL, Wise R, Rascol A *et al*. The anatomy of phonological and semantic processing in normal subjects. *Brain,* 115 (Pt 6): 1753-1768, 1992.

Desgranges B, Matuszewski V, Piolino P, Chetelat G, Mezenge F, Landeau B, De la Sayette V *et al*. Anatomical and functional alterations in semantic dementia: A voxel-based MRI and PET study. *Neurobiology of Aging,* 28: 1904-1913, 2007.

Drane DL, Ojemann GA, Aylward E, Ojemann JG, Johnson LC, Silbergeld DL, Miller JW *et al*. Category-specific naming and recognition deficits in temporal lobe epilepsy surgical patients. *Neuropsychologia,* 46: 1242-1255, 2008.

Dubois B, Feldman HH, Jacova C, Dekosky ST, Barberger-Gateau P, Cummings J, Delacourte A *et al*. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurology,* 6: 734-746, 2007.

Dunton J, Bruce C, Newton C. Investigating the impact of unfamiliar speaker accent on auditory comprehension in adults with aphasia. *International Journal of Language and Communication Disorders,* 46: 63-73, 2011.

Edwards-Lee T, Miller BL, Benson DF, Cummings JL, Russell GL, Boone K, Mena I. The temporal variant of frontotemporal dementia. *Brain,* 120 (Pt 6): 1027-1040, 1997.

Ellis AW, Young AW, Critchley EM. Loss of memory for people following temporal lobe damage. *Brain,* 112: 1469-1483, 1989.

Ellis HD, Jones DM, Mosdell N. Intra- and inter-modal repetition priming of familiar faces and voices. *British Journal of Psychology,* 88: 143-156, 1997.

Engel LR, Frum C, Puce A, Walker NA, Lewis JW. Different categories of living and non-living sound-sources activate distinct cortical networks. *Neuroimage,* 47: 1778-1791, 2009.

Eustache F, Lechevalier B, Viader F, Lambert J. Identification and discrimination disorders in auditory perception: a report on two cases. *Neuropsychologia,* 28: 257-270, 1990.

Eustache F, Lambert J, Cassier C, Dary M, Rossa Y, Rioux P, Viader F *et al*. Disorders of Auditory Identification in Dementia of the Alzheimer-Type. *Cortex,* 31: 119-127, 1995.

Evans BG, Iverson P. Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *Journal of the Acoustical Society of America,* 115: 352-361, 2004.

Evans JJ, Heggs AJ, Antoun N, Hodges JR. Progressive prosopagnosia associated with selective right temporal lobe atrophy. A new syndrome? *Brain,* 118: 1-13, 1995.

Fecteau S, Armony JL, Joanette Y, Belin P. Is voice processing species-specific in human auditory cortex? An fMRI study. *Neuroimage,* 23: 840-848, 2004.

Fecteau S, Belin P, Joanette Y, Armony JL. Amygdala responses to nonlinguistic emotional vocalizations. *Neuroimage,* 36: 480-487, 2007.

Fitch WT. The evolution of speech: a comparative review. *Trends in Cognitive Sciences,* 4: 258-267, 2000.

Flege JE, Munro MJ, MacKay IR. Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America,* 97: 3125-3134, 1995.

Floccia C, Goslin J, Girard F, Konopczynski G. Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance,* 32: 1276-1293, 2006.

Floccia C, Butler J, Goslin J, Ellis L. Regional and foreign accent processing in English: can listeners adapt? *Journal of Psycholinguistic Research,* 38: 379-412, 2009.

Francis AL, Nusbaum HC, Fenn K. Effects of training on the acoustic phonetic representation of synthetic speech. *Journal of Speech, Language, and Hearing Research,* 50: 1445-1465, 2007.

Friederici AD, Meyer M, von Cramon DY. Auditory language comprehension: An event-related fMRI study on the processing of syntactic and lexical information. *Brain and Language,* 74: 289-300, 2000.

Friederici AD, Kotz SA, Scott SK, Obleser J. Disentangling syntax and intelligibility in auditory language comprehension. *Human Brain Mapping,* 31: 448-457, 2010.

Gainotti G, Barbier A, Marra C. Slowly progressive defect in recognition of familiar people in a patient with right anterior temporal atrophy. *Brain,* 126: 792-803, 2003.

Gainotti G. Different patterns of famous people recognition disorders in patients with right and left anterior temporal lesions: a systematic review. *Neuropsychologia,* 45: 1591-1607, 2007a.

Gainotti G. Face familiarity feelings, the right temporal lobe and the possible underlying neural mechanisms. *Brain Research Reviews,* 56: 214-235, 2007b.

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,* 44: 238-248, 2008.

Gainotti G. What the study of voice recognition in normal subjects and brain-damaged patients tells us about models of familiar people recognition. *Neuropsychologia,* 49: 2273-2282, 2011.

Garrido L, Eisner F, McGettigan C, Stewart L, Sauter D, Hanley JR, Schweinberger SR *et al*. Developmental phonagnosia: a selective deficit of vocal identity recognition. *Neuropsychologia,* 47: 123-131, 2009.

Gates GA, Beiser A, Rees TS, D'Agostino RB, Wolf PA. Central auditory dysfunction may precede the onset of clinical dementia in people with probable Alzheimer's disease. *Journal of the American Geriatrics Society,* 50: 482-488, 2002.

Gentileschi V, Sperber S, Spinnler H. Progressive defective recognition of familiar people. *Neurocase,* 5: 407-424, 1999.

Gentileschi V, Sperber S, Spinnler H. Crossmodal agnosia for familiar people as a consequence of right infero-polar temporal atrophy. *Cognitive Neuropsychology,* 18: 439-463, 2001.

George MS, Parekh PI, Rosinsky N, Ketter TA, Kimbrell TA, Heilman KM, Herscovitch P *et al*. Understanding emotional prosody activates right hemisphere regions. *Archives of Neurology,* 53: 665-670, 1996.

Ghazanfar AA, Turesson HK, Maier JX, van Dinther R, Patterson RD, Logothetis NK. Vocal-tract resonances as indexical cues in rhesus monkeys. *Current Biology,* 17: 425-430, 2007.

Ghazanfar AA, Rendall D. Evolution of human vocal production. *Current Biology,* 18: R457-R460, 2008.

Giordano BL, McDonnell J, McAdams S. Hearing living symbols and nonliving icons: category specificities in the cognitive processing of environmental sounds. *Brain and Cognition,* 73: 7-19, 2010.

Giovanello KS, Alexander M, Verfaellie M. Differential impairment of person-specific knowledge in a patient with semantic dementia. *Neurocase,* 9: 15-26, 2003.

Giovannetti T, Sestito N, Libon DJ, Schmidt KS, Gallo JL, Gambino M, Chrysikou EG. The influence of personal familiarity on object naming, knowledge, and use in dementia. *Archives of Clinical Neuropsychology,* 21: 607-614, 2006.

Goldstone R. Influences of Categorization on Perceptual Discrimination. *Journal of Experimental Psychology-General,* 123: 178-200, 1994.

Goll JC, Crutch SJ, Warren JD. Central auditory disorders: toward a neuropsychology of auditory objects. *Current Opinion in Neurology,* 23: 617-627, 2010.

Goll JC, Crutch SJ, Loo JH, Rohrer JD, Frost C, Bamiou DE, Warren JD. Non-verbal sound processing in the primary progressive aphasias. *Brain,* 133: 272-285, 2010.

Goll JC, Kim LG, Ridgway GR, Hailstone JC, Lehmann M, Buckley AH, Crutch SJ *et al*. Impairments of auditory scene analysis in Alzheimer's disease. *Brain,* 135: 190-200, 2012.

Good CD, Johnsrude I, Ashburner J, Henson RN, Friston KJ, Frackowiak RS. Cerebral asymmetry and the effects of sex and handedness on brain structure: a voxel-based morphometric analysis of 465 normal adult human brains. *Neuroimage,* 14: 685-700, 2001.

Gorno-Tempini ML, Rankin KP, Woolley JD, Rosen HJ, Phengrasamy L, Miller BL. Cognitive and behavioral profile in a case of right anterior temporal lobe neurodegeneration. *Cortex,* 40: 631-644, 2004.

Gorno-Tempini ML, Dronkers NF, Rankin KP, Ogar JM, Phengrasamy L, Rosen HJ, Johnson JK *et al*. Cognition and anatomy in three variants of primary progressive aphasia. *Annals of Neurology,* 55: 335-346, 2004.

Gorno-Tempini ML, Brambati SM, Ginex V, Ogar J, Dronkers NF, Marcone A, Perani D *et al*. The logopenic/phonological variant of primary progressive aphasia. *Neurology,* 71: 1227-1234, 2008.

Gorno-Tempini ML, Hillis AE, Weintraub S, Kertesz A, Mendez M, Cappa SF, Ogar JM *et al*. Classification of primary progressive aphasia and its variants. *Neurology,* 76: 1006-1014, 2011.

Gosselin N, Peretz I, Johnsen E, Adolphs R. Amygdala damage impairs emotion recognition from music. *Neuropsychologia,* 45: 236-244, 2007.

Gosselin N, Peretz I, Hasboun D, Baulac M, Samson S. Impaired recognition of musical emotions and facial expressions following anteromedial temporal lobe excision. *Cortex,* 47: 1116-1125, 2011.

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. *Human Brain Mapping,* 13: 199-212, 2001.

Grainger J. Word-Frequency and Neighborhood Frequency-Effects in Lexical Decision and Naming. *Journal of Memory and Language,* 29: 228-244, 1990.

Greene JD, Hodges JR. Identification of famous faces and famous names in early Alzheimer's disease. Relationship to anterograde episodic and general semantic memory. *Brain,* 119 (Pt 1): 111-128, 1996.

Griffiths TD, Rees A, Witton C, Cross PM, Shakir RA, Green GGR. Spatial and temporal auditory processing deficits following right hemisphere infarction - A psychophysical study. *Brain,* 120: 785-794, 1997.

Griffiths TD, Buchel C, Frackowiak RS, Patterson RD. Analysis of temporal structure in sound by the human brain. *Nature Neuroscience,* 1: 422-427, 1998.

Griffiths TD, Rees A, Green GGR. Disorders of human complex sound processing. *Neurocase,* 5: 365-378, 1999.

Griffiths TD, Warren JD. The planum temporale as a computational hub. *Trends in Neurosciences,* 25: 348-353, 2002.

Griffiths TD, Warren JD. What is an auditory object? *Nature Reviews Neuroscience,* 5: 887-892, 2004.

Griffiths TD, Kumar S, Warren JD, Stewart L, Stephan KE, Friston KJ. Approaches to the cortical analysis of auditory objects. *Hearing Research,* 229: 46-53, 2007.

Grossman M, Mickanin J, Onishi K, Hughes E, DEsposito M, Ding XS, Alavi A *et al*. Progressive nonfluent aphasia: Language, cognitive, and PET measures contrasted with probable Alzheimer's disease. *Journal of Cognitive Neuroscience,* 8: 135-154, 1996.

Grossman M, Ash S. Primary progressive aphasia: A review. *Neurocase,* 10: 3-18, 2004.

Gunawardena D, Ash S, McMillan C, Avants B, Gee J, Grossman M. Why are patients with progressive nonfluent aphasia nonfluent? *Neurology,* 75: 588-594, 2010.

Hailstone JC, Omar R, Warren JD. Relatively preserved knowledge of music in semantic dementia. *Journal of Neurology, Neurosurgery and Psychiatry,* 80: 808-809, 2009.

Hailstone JC, Omar R, Henley SM, Frost C, Kenward MG, Warren JD. It's not what you play, it's how you play it: timbre affects perception of emotion in music. *Quarterly Journal of Experimental Psychology,* 62: 2141-2155, 2009.

Hall DA, Anderson CA, Filley CM, Newcombe J, Hughes RL. A French accent after corpus callosum infarct. *Neurology,* 60: 1551-1552, 2003.

Halpern AR, Bartlett JC, Dowling WJ. Aging and experience in the recognition of musical transpositions. *Psychology and Aging,* 10: 325-342, 1995.

Halpern AR, Zatorre RJ, Bouffard M, Johnson JA. Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia,* 42: 1281-1292, 2004.

Hanley JR, Young AW, Pearson NA. Defective recognition of familiar people. *Cognitive Neuropsychology,* 6: 179-210, 1989.

Hanley JR, Smith ST, Hadfield J. I recognise you but I can't place you: an investigation of familiar-only experiences during tests of voice and face recognition. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology,* 51: 179-195, 1998.

Hanley JR, Turner JM. Why are familiar-only experiences more frequent for voices than for faces? *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology,* 53: 1105-1116, 2000.

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,* 17: 830-839, 2009.

Hanson HM. Glottal characteristics of female speakers: acoustic correlates. *Journal of the Acoustical Society of America,* 101: 466-481, 1997.

Heffner HE, Heffner RS. Hearing loss in Japanese macaques following bilateral auditory cortex lesions. *Journal of Neurophysiology,* 55: 256-271, 1986.

Hellstrom A, Almkvist O. Tone duration discrimination in demented, memory-impaired, and healthy elderly. *Dementia and Geriatric Cognitive Disorders,* 8: 49-54, 1997.

Henley SM, Wild EJ, Hobbs NZ, Scahill RI, Ridgway GR, Macmanus DG, Barker RA *et al*. Relationship between CAG repeat length and brain volume in premanifest and early Huntington's disease. *Journal of Neurology,* 256: 203-212, 2009.

Hillenbrand JM, Houde RA. A narrow band pattern-matching model of vowel perception. *Journal of the Acoustical Society of America,* 113: 1044-1055, 2003.

Hodges JR, Salmon DP, Butters N. Semantic Memory Impairment in Alzheimers-Disease - Failure of Access Or Degraded Knowledge. *Neuropsychologia,* 30: 301-314, 1992.

Horley K, Reid A, Burnham D. Emotional prosody perception and production in dementia of the Alzheimer's type. *Journal of Speech, Language, and Hearing Research,* 53: 1132-1146, 2010.

Howell P, Barry W, Vinson D. Strength of British English accents in altered listening conditions. *Perception & Psychophysics,* 68: 139-153, 2006.

Humphreys GW, Riddoch MJ. On the Case for Multiple Semantic Systems - A Reply. *Cognitive Neuropsychology,* 5: 143-150, 1988.

Humphries C, Willard K, Buchsbaum B, Hickok G. Role of anterior temporal cortex in auditory sentence comprehension: an fMRI study. *Neuroreport,* 12: 1749-1752, 2001.

Imaizumi S, Mori K, Kiritani S, Kawashima R, Sugiura M, Fukuda H, Itoh K *et al*. Vocal identification of speaker and emotion activates different brain regions. *Neuroreport,* 8: 2809-2812, 1997.

Iverson P, Kuhl PK. Perceptual magnet and phoneme boundary effects in speech perception: do they arise from a common mechanism? *Perception & Psychophysics,* 62: 874-886, 2000.

Iverson P, Kuhl PK, Akahane-Yamada R, Diesch E, Tohkura Y, Kettermann A, Siebert C. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition,* 87: B47-B57, 2003.

Ives DT, Smith DRR, Patterson RD. Discrimination of speaker size from syllable phrases. *Journal of the Acoustical Society of America,* 118: 3816-3822, 2005.

Jackson M, Warrington EK. Arithmetic skills in patients with unilateral cerebral lesions. *Cortex,* 22: 611-620, 1986.

Jancke L, Wustenberg T, Scheich H, Heinze HJ. Phonetic perception and the temporal cortex. *Neuroimage,* 15: 733-746, 2002.

Jorgens S, Biermann-Ruben K, Kurz MW, Flugel C, Kurz KD, Antke C, Hartung HP *et al*. Word deafness as a cortical auditory processing deficit: A case report with MEG. *Neurocase,* 14: 307-316, 2008.

Josephs KA, Whitwell JL, Vemuri P, Senjem ML, Boeve BF, Knopman DS, Smith GE *et al*. The anatomic correlate of prosopagnosia in semantic dementia. *Neurology,* 71: 1628-1633, 2008.

Josephs KA, Hodges JR, Snowden JS, Mackenzie IR, Neumann M, Mann DM, Dickson DW. Neuropathological background of phenotypical variability in frontotemporal dementia. *Acta Neuropathologica,* 122: 137-152, 2011.

Joubert S, Felician O, Barbeau E, Sontheimer A, Barton JJ, Ceccaldi M, Poncet M. Impaired configurational processing in a case of progressive prosopagnosia associated with predominant right temporal lobe atrophy. *Brain,* 126: 2537-2550, 2003.

Joubert S, Mauries S, Barbeau E, Ceccaldi M, Poncet M. The role of context in remembering familar persons: insights from semantic dementia. *Brain and Cognition,* 55: 254-261, 2004.

Joubert S, Felician O, Barbeau E, Sontheimer A, Guedj E, Ceccaldi M, Poncet M. Progressive prosopagnosia: clinical and neuroimaging results. *Neurology,* 63: 1962-1965, 2004.

Joubert S, Felician O, Barbeau E, Ranjeva JP, Christophe M, Didic M, Poncet M *et al*. The right temporal lobe variant of frontotemporal dementia: cognitive and neuroanatomical profile of three patients. *Journal of Neurology,* 253: 1447-1458, 2006.

Juslin PN, Laukka P. Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion,* 1: 381-412, 2001.

Juslin PN, Laukka P. Communication of emotions in vocal expression and music performance: different channels, same code? *Psychological Bulletin,* 129: 770-814, 2003.

Kaas JH, Hackett TA. 'What' and 'where' processing in auditory cortex. *Nature Neuroscience,* 2: 1045-1047, 1999.

Kanwisher N, McDermott J, Chun MM. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience,* 17: 4302-4311, 1997.

Karpf A. The human voice: the story of a remarkable talent. London: Bloomsbury Publishing; 2006.

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.

Kay J, Lesser R, Coltheart M. PALPA: psycholinguistic assessments of language processing in aphasia. London: Psychology Press; 1992.

Keane J, Calder AJ, Hodges JR, Young AW. Face and emotion processing in frontal variant frontotemporal dementia. *Neuropsychologia,* 40: 655-665, 2002.

Kipps CM, Nestor PJ, Acosta-Cabronero J, Arnold R, Hodges JR. Understanding social dysfunction in the behavioural variant of frontotemporal dementia: the role of emotion and sarcasm processing. *Brain,* 132: 592-603, 2009.

Kisilevsky BS, Hains SMJ, Lee K, Xie X, Huang HF, Ye HH, Zhang K *et al*. Effects of experience on fetal voice recognition. *Psychological Science,* 14: 220-224, 2003.

Klatt DH, Klatt LC. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America,* 87: 820-857, 1990.

Knosche TR, Lattner S, Maess B, Schauer M, Friederici AD. Early parallel processing of auditory word and voice information. *Neuroimage,* 17: 1493-1503, 2002.

Kopelman MD, Stanhope N, Kingsley D. Retrograde amnesia in patients with diencephalic, temporal lobe or frontal lesions. *Neuropsychologia,* 37: 939-958, 1999.

Kurowski KM, Blumstein SE, Alexander M. The foreign accent syndrome: a reconsideration. *Brain and Language,* 54: 1-25, 1996.

Kurylo DD, Corkin S, Allard T, Zatorre RJ, Growdon JH. Auditory Function in Alzheimers-Disease. *Neurology,* 43: 1893-1899, 1993.

Lambon Ralph MA, Graham KS, Ellis AW, Hodges JR. Naming in semantic dementia--what matters? *Neuropsychologia,* 36: 775-784, 1998.

Lambon Ralph MA, Graham KS, Patterson K, Hodges JR. Is a picture worth a thousand words? Evidence from concept definitions by patients with semantic dementia. *Brain and Language,* 70: 309-335, 1999.

Lambon Ralph MA, McClelland JL, Patterson K, Galton CJ, Hodges JR. No right to speak? The relationship between object naming and semantic impairment: neuropsychological evidence and a computational model. *Journal of Cognitive Neuroscience,* 13: 341-356, 2001.

Lambon Ralph MA, Patterson K, Garrard P, Hodges JR. Semantic dementia with category specificity: A comparative case-series study. *Cognitive Neuropsychology,* 20: 307-326, 2003.

Lambon Ralph MA, Patterson K, Graham N, Dawson K, Hodges JR. Homogeneity and heterogeneity in mild cognitive impairment and Alzheimer's disease: a cross-sectional and longitudinal study of 55 cases. *Brain,* 126: 2350-2362, 2003.

Lambon Ralph MA, Patterson K. Generalization and differentiation in semantic memory: insights from semantic dementia. *Annals of the New York Academy of Sciences,* 1124: 61-76, 2008.

Lang CJ, Kneidl O, Hielscher-Fastabend M, Heckmann JG. Voice recognition in aphasic and non-aphasic stroke patients. *Journal of Neurology,* 256: 1303-1306, 2009.

Lattner S, Meyer ME, Friederici AD. Voice perception: Sex, pitch, and the right hemisphere. *Human Brain Mapping,* 24: 11-20, 2005.

Lauro-Grotto R, Piccini C, Shallice T. Modality-specific operations in semantic dementia. *Cortex,* 33: 593-622, 1997.

Lavner Y, Gath I, Rosenhouse J. The effects of acoustic modifications on the identification of familiar voices speaking isolated vowels. *Speech Communication,* 30: 9-26, 2000.

Leaver AM, Rauschecker JP. Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *Journal of Neuroscience,* 30: 7604-7612, 2010.

Lee AC, Buckley MJ, Pegman SJ, Spiers H, Scahill VL, Gaffan D, Bussey TJ *et al*. Specialization in the medial temporal lobe for processing of objects and scenes. *Hippocampus,* 15: 782-797, 2005.

Lee ACH, Buckley MJ, Gaffan D, Emery T, Hodges JR, Graham KS. Differentiating the roles of the hippocampus and perirhinal cortex in processes beyond long-term declarative memory: A double dissociation in dementia. *Journal of Neuroscience,* 26: 5198-5203, 2006.

Leff AP, Iverson P, Schofield TM, Kilner JM, Crinion JT, Friston KJ, Price CJ. Vowel-specific mismatch responses in the anterior superior temporal gyrus: an fMRI study. *Cortex,* 45: 517-526, 2009.

Lehmann M, Crutch SJ, Ridgway GR, Ridha BH, Barnes J, Warrington EK, Rossor MN *et al*. Cortical thickness and voxel-based morphometry in posterior cortical atrophy and typical Alzheimer's disease. *Neurobiology of Aging,* 2009.

Lehmann M, Barnes J, Ridgway GR, Wattam-Bell J, Warrington EK, Fox NC, Crutch SJ. Basic Visual Function and Cortical Thickness Patterns in Posterior Cortical Atrophy. *Cerebral Cortex,* 2011.

Levy DA, Granot R, Bentin S. Processing specificity for human voice stimuli: electrophysiological evidence. *Neuroreport,* 12: 2653-2657, 2001.

Lewis JW, Wightman FL, Brefczynski JA, Phinney RE, Binder JR, DeYoe EA. Human brain regions involved in recognizing environmental sounds. *Cerebral Cortex,* 14: 1008-1021, 2004.

Lewis JW, Brefczynski JA, Phinney RE, Janik JJ, DeYoe EA. Distinct cortical pathways for processing tool versus animal sounds. *Journal of Neuroscience,* 25: 5148-5158, 2005.

Lewis JW, Talkington WJ, Walker NA, Spirou GA, Jajosky A, Frum C, Brefczynski-Lewis JA. Human cortical organization for processing vocalizations indicates representation of harmonic structure as a signal attribute. *Journal of Neuroscience,* 29: 2283-2296, 2009.

Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA. Neural substrates of phonemic perception. *Cerebral Cortex,* 15: 1621-1631, 2005.

Lucchelli F, Spinnler H. A reappraisal of person recognition and identification. *Cortex,* 44: 230-237, 2008.

Luce PA, Pisoni DB. Recognizing spoken words: the neighborhood activation model. *Ear and Hearing,* 19: 1-36, 1998.

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,* 45: 1823-1831, 2007.

Luzzi S, Viticchi G, Piccirilli M, Fabi K, Pesallaccia M, Bartolini M, Provinciali L *et al*. Foreign accent syndrome as the initial sign of primary progressive aphasia. *Journal of Neurology, Neurosurgery and Psychiatry,* 79: 79-81, 2008.

Lyons F, Kay J, Hanley JR, Haslam C. Selective preservation of memory for people in the context of semantic memory disorder: patterns of association and dissociation. *Neuropsychologia,* 44: 2887-2898, 2006.

Mahon BZ, Anzellotti S, Schwarzbach J, Zampini M, Caramazza A. Category-specific organization in the human brain does not require visual experience. *Neuron,* 63: 397-405, 2009.

Martin A, Chao LL. Semantic memory and the brain: structure and processes. *Current Opinion in Neurobiology,* 11: 194-201, 2001.

Martin A. The representation of object concepts in the brain. *Annual Review of Psychology,* 58: 25-45, 2007.

Masataka N. Development of Vocal Recognition of Mothers in Infant Japanese Macaques. *Developmental Psychobiology,* 18: 107-114, 1985.

Mazzucchi A, Marchini C, Budai R, Parma M. A case of receptive amusia with prominent timbre perception defect. *Journal of Neurology, Neurosurgery and Psychiatry,* 45: 644-647, 1982.

McCarthy RA, Warrington EK. Evidence for modality-specific meaning systems in the brain. *Nature,* 334: 428-430, 1988.

McCarthy RA, Warrington EK. Actors but not scripts: the dissociation of people and events in retrograde amnesia. *Neuropsychologia,* 30: 633-644, 1992.

McNemar QR. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika,* 12: 153-157, 1947.

Menon V, Levitin DJ, Smith BK, Lembke A, Krasnow BD, Glazer D, Glover GH *et al*. Neural correlates of timbre change in harmonic sounds. *Neuroimage,* 17: 1742-1754, 2002.

Mesulam M, Rogalski E, Wieneke C, Cobia D, Rademaker A, Thompson C, Weintraub S. Neurology of anomia in the semantic variant of primary progressive aphasia. *Brain,* 132: 2553-2565, 2009.

Mesulam MM. Primary progressive aphasia. *Annals of Neurology,* 49: 425-432, 2001.

Meyer M, Alter K, Friederici AD, Lohmann G, von Cramon DY. FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Human Brain Mapping,* 17: 73-88, 2002.

Meyer M, Steinhauer K, Alter K, Friederici AD, von Cramon DY. Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain and Language,* 89: 277-289, 2004.

Miller BL, Chang L, Mena I, Boone K, Lesser IM. Progressive right frontotemporal degeneration: clinical, neuropsychological and SPECT characteristics. *Dementia,* 4: 204-213, 1993.

Minoshima S, Giordani B, Berent S, Frey KA, Foster NL, Kuhl DE. Metabolic reduction in the posterior cingulate cortex in very early Alzheimer's disease. *Annals of Neurology,* 42: 85-94, 1997.

Mion M, Patterson K, Acosta-Cabronero J, Pengas G, Izquierdo-Garcia D, Hong YT, Fryer TD *et al*. What the left and right anterior fusiform gyri tell us about semantic memory. *Brain,* 133: 3256-3268, 2010.

Mitchell RL, Elliott R, Barry M, Cruttenden A, Woodruff PW. The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia,* 41: 1410-1421, 2003.

Nakamura K, Kawashima R, Sugiura M, Kato T, Nakamura A, Hatano K, Nagumo S *et al*. Neural substrates for recognition of familiar voices: a PET study. *Neuropsychologia,* 39: 1047-1054, 2001.

Nathan L, Wells B, Donlan C. Children's comprehension of unfamiliar regional accents: a preliminary investigation. *Journal of Child Language,* 25: 343-365, 1998.

Nearey TM. Speech perception as pattern recognition. *Journal of the Acoustical Society of America,* 101: 3241-3254, 1997.

Neary D, Snowden JS, Shields RA, Burjan AW, Northen B, MacDermott N, Prescott MC *et al*. Single photon emission tomography using 99mTc-HM-PAO in the investigation of dementia. *Journal of Neurology, Neurosurgery and Psychiatry,* 50: 1101-1109, 1987.

Neary D, Snowden JS, Gustafson L, Passant U, Stuss D, Black S, Freedman M *et al*. Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. *Neurology,* 51: 1546-1554, 1998.

Neary D, Snowden JS, Mann DM. Classification and description of frontotemporal dementias. *Annals of the New York Academy of Sciences,* 920: 46-51, 2000.

Nelson HE. National Adult Reading Test (NART): Test manual. Windsor: NFER Nelson; 1982.

Nestor PJ, Fryer TD, Hodges JR. Declarative memory impairments in Alzheimer's disease and semantic dementia. *Neuroimage,* 30: 1010-1020, 2006.

Neuner F, Schweinberger SR. Neuropsychological impairments in the recognition of faces, voices, and personal names. *Brain and Cognition,* 44: 342-366, 2000.

Noonan KA, Jefferies E, Corbett F, Lambon Ralph MA. Elucidating the nature of deregulated semantic cognition in semantic aphasia: evidence for the roles of prefrontal and temporo-parietal cortices. *Journal of Cognitive Neuroscience,* 22: 1597-1613, 2010.

Norris D, McQueen JM, Cutler A. Perceptual learning in speech. *Cognitive Psychology,* 47: 204-238, 2003.

Nygaard LC, Sommers MS, Pisoni DB. Speech perception as a talker-contingent process. *Psychological Science,* 5: 42-46, 1994.

Nygaard LC, Pisoni DB. Talker-specific learning in speech perception. *Perception & Psychophysics,* 60: 355-376, 1998.

Ockleford EM, Vince MA, Layton C, Reader MR. Responses of Neonates to Parents and Others Voices. *Early Human Development,* 18: 27-36, 1988.

Olson IR, Plotzker A, Ezzyat Y. The Enigmatic temporal pole: a review of findings on social and emotional processing. *Brain,* 130: 1718-1731, 2007.

Omar R, Rohrer JD, Hailstone JC, Warren JD. Structural neuroanatomy of face processing in frontotemporal lobar degeneration. *Journal of Neurology, Neurosurgery and Psychiatry,* 2010.

Omar R, Hailstone JC, Warren JE, Crutch SJ, Warren JD. The cognitive organization of music knowledge: a clinical analysis. *Brain,* 133: 1200-1213, 2010.

Omar R, Henley SM, Bartlett JW, Hailstone JC, Gordon E, Sauter DA, Frost C *et al*. The structural neuroanatomy of music emotion recognition: evidence from frontotemporal lobar degeneration. *Neuroimage,* 56: 1814-1821, 2011.

Otsuki M, Soma Y, Sato M, Homma A, Tsuji S. Slowly progressive pure word deafness. *European Neurology,* 39: 135-140, 1998.

Papagno C, Capitani E. Proper name anomia: a case with sparing of the first-letter knowledge. *Neuropsychologia,* 36: 669-679, 1998.

Papcun G, Kreiman J, Davis A. Long-Term-Memory for Unfamiliar Voices. *Journal of the Acoustical Society of America,* 85: 913-925, 1989.

Patterson ML, Werker JF. Infants' ability to match dynamic phonetic and gender, information in the face and voice. *Journal of Experimental Child Psychology,* 81: 93-115, 2002.

Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD. The processing of temporal pitch and melody information in auditory cortex. *Neuron,* 36: 767-776, 2002.

Paulesu E, Frith CD, Frackowiak RS. The neural correlates of the verbal component of working memory. *Nature,* 362: 342-345, 1993.

Peretz I, Kolinsky R, Tramo M, Labrecque R, Hublet C, Demeurisse G, Belleville S. Functional dissociations following bilateral lesions of auditory cortex. *Brain,* 117: 1283-1301, 1994.

Peretz I, Zatorre RJ. Brain organization for music processing. *Annual Review of Psychology,* 56: 89-114, 2005.

Perrachione TK, Wong PC. Learning to recognize speakers of a non-native language: implications for the functional organization of human auditory cortex. *Neuropsychologia,* 45: 1899-1910, 2007.

Perry RJ, Hodges JR. Relationship between functional and neuropsychological performance in early Alzheimer disease. *Alzheimer Disease and Associated Disorders,* 14: 1-10, 2000.

Perry RJ, Rosen HR, Kramer JH, Beer JS, Levenson RL, Miller BL. Hemispheric dominance for emotions, empathy and social behaviour: evidence from right and left handers with frontotemporal dementia. *Neurocase,* 7: 145-160, 2001.

Peschke C, Ziegler W, Eisenberger J, Baumgaertner A. Phonological manipulation between speech perception and production activates a parieto-frontal circuit. *Neuroimage,* 59: 788-799, 2012.

Petkov CI, Kayser C, Steudel T, Whittingstall K, Augath M, Logothetis NK. A voice region in the monkey brain. *Nature Neuroscience,* 11: 367-374, 2008.

Piguet O, Hornberger M, Mioshi E, Hodges JR. Behavioural-variant frontotemporal dementia: diagnosis, clinical staging, and management. *Lancet Neurology,* 10: 162-172, 2011.

Piwnica-Worms KE, Omar R, Hailstone JC, Warren JD. Flavour processing in semantic dementia. *Cortex,* 46: 761-768, 2010.

Pollack I, Pickett JM, Sumby WH. On the Identification of Speakers by Voice. *Journal of the Acoustical Society of America,* 26: 403-406, 1954.

Rami L, Loy CT, Hailstone J, Warren JD. Odour identification in frontotemporal lobar degeneration. *Journal of Neurology,* 254: 431-435, 2007.

Rankin KP, Salazar A, Gorno-Tempini ML, Sollberger M, Wilson SM, Pavlic D, Stanley CM *et al*. Detecting sarcasm from paralinguistic cues: anatomic and cognitive correlates in neurodegenerative disease. *Neuroimage,* 47: 2005-2015, 2009.

Rapcsak SZ, Kentros M, Rubens AB. Impaired Recognition of Meaningful Sounds in Alzheimers-Disease. *Archives of Neurology,* 46: 1298-1300, 1989.

Rascovsky K, Hodges JR, Kipps CM, Johnson JK, Seeley WW, Mendez MF, Knopman D *et al*. Diagnostic criteria for the behavioral variant of frontotemporal dementia (bvFTD): current limitations and future directions. *Alzheimer Disease and Associated Disorders,* 21: S14-S18, 2007.

Rascovsky K, Hodges JR, Knopman D, Mendez MF, Kramer JH, Neuhaus J, van Swieten JC *et al*. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain,* 134: 2456-2477, 2011.

Rauschecker JP. Parallel processing in the auditory cortex of primates. *Audiology and Neuro-Otology,* 3: 86-103, 1998.

Rauschecker JP, Scott SK. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience,* 12: 718-724, 2009.

Reichman WE, Negron A. Negative symptoms in the elderly patient with dementia. *International Journal of Geriatric Psychiatry,* 16 Suppl 1: S7-11, 2001.

Reilly J, Peelle JE, Antonucci SM, Grossman M. Anomia as a marker of distinct semantic memory impairments in Alzheimer's disease and semantic dementia. *Neuropsychology,* 25: 413-426, 2011.

Remedios R, Logothetis NK, Kayser C. An auditory region in the primate insular cortex responding preferentially to vocal communication sounds. *Journal of Neuroscience,* 29: 1034-1045, 2009.

Remez RE, Fellowes JM, Rubin PE. Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance,* 23: 651-666, 1997.

Rendall D. Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons. *Journal of the Acoustical Society of America,* 113: 3390-3402, 2003.

Riddoch MJ, Humphreys GW. A case of integrative visual agnosia. *Brain,* 110 (Pt 6): 1431-1462, 1987.

Riddoch MJ, Humphreys GW. Visual agnosia. *Neurologic Clinics,* 21: 501-520, 2003.

Ridgway GR, Omar R, Ourselin S, Hill DL, Warren JD, Fox NC. Issues with threshold masking in voxel-based morphometry of atrophied brains. *Neuroimage,* 44: 99-111, 2009.

Ridha BH, Rossor MN. A cognitive bedside assessment beyond the MMSE: the Addenbrooke's Cognitive Examination. *Practical Neurology,* 7: 245-249, 2007.

Roberts VJ, Ingram SM, Lamar M, Green RC. Prosody impairment and associated affective and behavioral disturbances in Alzheimer's disease. *Neurology,* 47: 1482-1488, 1996.

Rogers TT, Ivanoiu A, Patterson K, Hodges JR. Semantic memory in Alzheimer's disease and the frontotemporal dementias: a longitudinal study of 236 patients. *Neuropsychology,* 20: 319-335, 2006.

Rohrer JD, Sauter D, Scott S, Rossor MN, Warren JD. Receptive prosody in nonfluent primary progressive aphasias. *Cortex,* 48: 308-316, 2010.

Rohrer JD, Ridgway GR, Modat M, Ourselin S, Mead S, Fox NC, Rossor MN *et al*. Distinct profiles of brain atrophy in frontotemporal lobar degeneration caused by progranulin and tau mutations. *Neuroimage,* 53: 1070-1076, 2010.

Rohrer JD, Schott JM. Primary Progressive Aphasia - Defining Genetic and Pathological Subtypes. *Current Alzheimer Research,* 8: 266-272, 2011.

Rohrer JD, Rossor MN, Warren JD. Alzheimer's pathology in primary progressive aphasia. *Neurobiology of Aging,* 33: 744-752, 2012.

Rosen HJ, Perry RJ, Murphy J, Kramer JH, Mychack P, Schuff N, Weiner M *et al*. Emotion comprehension in the temporal variant of frontotemporal dementia. *Brain,* 125: 2286-2295, 2002.

Rosen HJ, Wilson MR, Schauer GF, Allison S, Gorno-Tempini ML, Pace-Savitsky C, Kramer JH *et al*. Neuroanatomical correlates of impaired recognition of emotion in dementia. *Neuropsychologia,* 44: 365-373, 2006.

Ross ED, Monnot M. Neurology of affective prosody and its functional-anatomic organization in right hemisphere. *Brain and Language,* 104: 51-74, 2008.

Samson S, Zatorre RJ. Learning and retention of melodic and verbal information after unilateral temporal lobectomy. *Neuropsychologia,* 30: 815-826, 1992.

Saslove H, Yarmey AD. Long-term auditory memory: speaker identification. *Journal of Applied Psychology,* 65: 111-116, 1980.

Sauter DA, Scott SK. More than one kind of happiness: Can we recognize vocal expression of different positive states? *Motivation and Emotion,* 31: 192-199, 2007.

Sauter DA, Calder AJ, Eisner F, Scott SK. Perceptual cues in non-verbal vocal expressions of emotion. *Quarterly Journal of Experimental Psychology,* 63: 2251-2272, 2010.

Saygin AP, Leech R, Dick F. Nonverbal auditory agnosia with lesion to Wernicke's area. *Neuropsychologia,* 48: 107-113, 2010.

Scahill RI, Schott JM, Stevens JM, Rossor MN, Fox NC. Mapping the evolution of regional atrophy in Alzheimer's disease: unbiased analysis of fluid-registered serial MRI. *Proceedings of the National Academy of Sciences of the United States of America,* 99: 4703-4707, 2002.

Scharinger M, Monahan PJ, Idsardi WJ. You had me at "Hello": rapid extraction of dialect information from spoken words. *Neuroimage,* 15: 2329-2338, 2011.

Schmale R, Seidl A. Accommodating variability in voice and foreign accent: flexibility of early word representations. *Developmental Science,* 12: 583-601, 2009.

Schneider P, Sluming V, Roberts N, Scherg M, Goebel R, Specht HJ, Dosch HG *et al*. Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. *Nature Neuroscience,* 8: 1241-1247, 2005.

Schonell F, Goodacre E. The psychology and teaching of reading. London: Oliver Boyd; 1971.

Schonwiesner M, Novitski N, Pakarinen S, Carlson S, Tervaniemi M, Naatanen R. Heschl's gyrus, posterior superior temporal gyrus, and mid-ventrolateral prefrontal cortex have different roles in the detection of acoustic changes. *Journal of Neurophysiology,* 97: 2075-2082, 2007.

Schweinberger SR, Herholz A, Sommer W. Recognizing famous voices: influence of stimulus duration and different types of retrieval cues. *Journal of Speech and Hearing Research,* 40: 453-463, 1997.

Schweinberger SR, Herholz A, Stief V. Auditory long-term memory: repetition priming of voice recognition. *Quarterly Journal of Experimental Psychology,* 50: 498-517, 1997.

Schweinberger SR. Human brain potential correlates of voice priming and voice recognition. *Neuropsychologia,* 39: 921-936, 2001.

Schweinberger SR, Burton AM. Covert recognition and the neural system for face processing. *Cortex,* 39: 9-30, 2003.

Scott SK, Blank CC, Rosen S, Wise RJ. Identification of a pathway for intelligible speech in the left temporal lobe. *Brain,* 123 Pt 12: 2400-2406, 2000.

Scott SK, Rosen S, Lang H, Wise RJ. Neural correlates of intelligibility in speech investigated with noise vocoded speech--a positron emission tomography study. *Journal of the Acoustical Society of America,* 120: 1075-1083, 2006.

Seeley WW, Menon V, Schatzberg AF, Keller J, Glover GH, Kenna H, Reiss AL *et al*. Dissociable intrinsic connectivity networks for salience processing and executive control. *Journal of Neuroscience,* 27: 2349-2356, 2007.

Seeley WW, Crawford R, Rascovsky K, Kramer JH, Weiner M, Miller BL, Gorno-Tempini ML. Frontal paralimbic network atrophy in very mild behavioral variant frontotemporal dementia. *Archives of Neurology,* 65: 249-255, 2008.

Seeley WW, Crawford RK, Zhou J, Miller BL, Greicius MD. Neurodegenerative diseases target large-scale human brain networks. *Neuron,* 62: 42-52, 2009.

Shah NJ, Marshall JC, Zafiris O, Schwab A, Zilles K, Markowitsch HJ, Fink GR. The neural correlates of person familiarity. A functional magnetic resonance imaging study with clinical implications. *Brain,* 124: 804-815, 2001.

Sharp DJ, Scott SK, Wise RJ. Retrieving meaning after temporal lobe infarction: the role of the basal language area. *Annals of Neurology,* 56: 836-846, 2004.

Singh NC, Theunissen FE. Modulation spectra of natural sounds and ethological theories of auditory processing. *Journal of the Acoustical Society of America,* 114: 3394-3411, 2003.

Singh S, Murry T. Multidimensional classification of normal voice qualities. *Journal of the Acoustical Society of America,* 64: 81-87, 1978.

Smith DR, Patterson RD. The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *Journal of the Acoustical Society of America,* 118: 3177-3186, 2005.

Smith DR, Patterson RD, Turner R, Kawahara H, Irino T. The processing and perception of size information in speech sounds. *Journal of the Acoustical Society of America,* 117: 305-318, 2005.

Smith DR, Walters TC, Patterson RD. Discrimination of speaker sex and size when glottal-pulse rate and vocal-tract length are controlled. *Journal of the Acoustical Society of America,* 122: 3628-3639, 2007.

Snowden JS, Griffiths HL, Neary D. Semantic dementia: autobiographical contribution to preservation of meaning. *Cognitive Neuropsychology,* 11: 265-288, 1994.

Snowden JS, Bathgate D, Varma A, Blackshaw A, Gibbons ZC, Neary D. Distinct behavioural profiles in frontotemporal dementia and semantic dementia. *Journal of Neurology, Neurosurgery and Psychiatry,* 70: 323-332, 2001.

Snowden JS, Thompson JC, Neary D. Knowledge of famous faces and names in semantic dementia. *Brain,* 127: 860-872, 2004.

Snowden JS, Austin NA, Sembi S, Thompson JC, Craufurd D, Neary D. Emotion recognition in Huntington's disease and frontotemporal dementia. *Neuropsychologia,* 46: 2638-2649, 2008.

Sokhi DS, Hunter MD, Wilkinson ID, Woodruff PW. Male and female voices activate distinct regions in the male brain. *Neuroimage,* 27: 572-578, 2005.

Sonty SP, Mesulam MM, Weintraub S, Johnson NA, Parrish TB, Gitelman DR. Altered effective connectivity within the language network in primary progressive aphasia. *Journal of Neuroscience,* 27: 1334-1345, 2007.

Squire LR, Stark CE, Clark RE. The medial temporal lobe. *Annual Review of Neuroscience,* 27: 279-306, 2004.

Staeren N, Renvall H, De MF, Goebel R, Formisano E. Sound categories are represented as distributed patterns in the human auditory cortex. *Current Biology,* 19: 498-502, 2009.

Stopford CL, Snowden JS, Thompson JC, Neary D. Distinct memory profiles in Alzheimer's disease. *Cortex,* 43: 846-857, 2007.

Stopford CL, Snowden JS, Thompson JC, Neary D. Variability in cognitive presentation of Alzheimer's disease. *Cortex,* 44: 185-195, 2008.

Taler V, Baum SR, Chertkow H, Saumier D. Comprehension of grammatical and emotional prosody is impaired in Alzheimer's disease. *Neuropsychology,* 22: 188-195, 2008.

Tanji K, Suzuki K, Okuda J, Shimizu H, Seki H, Kimura I, Endo K *et al*. Formant interaction as a cue to vowel perception: a case report. *Neurocase,* 9: 350-355, 2003.

Taylor A, Warrington EK. Visual agnosia: a single case report. *Cortex,* 7: 152-161, 1971.

Testa JA, Beatty WW, Gleason AC, Orbelo DM, Ross ED. Impaired affective prosody in AD: relationship to aphasic deficits and emotional behaviors. *Neurology,* 57: 1474-1481, 2001.

Thierry G, Giraud AL, Price C. Hemispheric dissociation in access to the human semantic system. *Neuron,* 38: 499-506, 2003.

Thioux M, Pillon A, Samson D, de Partz MP, Noel MP, Seron X. The isolation of numerals at the semantic level. *Neurocase,* 4: 371-389, 1998.

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,* 42: 359-370, 2004.

Titze IR. The human instrument. *Scientific American,* 298: 94-101, 2008.

Toivonen M, Rama P. N400 during recognition of voice identity and vocal affect. *Neuroreport,* 20: 1245-1249, 2009.

Tranel D, Damasio H, Damasio AR. A neural basis for the retrieval of conceptual knowledge. *Neuropsychologia,* 35: 1319-1327, 1997.

Tranel D. Impaired naming of unique landmarks is associated with left temporal polar damage. *Neuropsychology,* 20: 1-10, 2006.

Tucker DM, Watson RT, Heilman KM. Discrimination and evocation of affectively intoned speech in patients with right parietal disease. *Neurology,* 27: 947-950, 1977.

Turkeltaub PE, Coslett HB. Localization of sublexical speech perception components. *Brain and Language,* 114: 1-15, 2010.

Uttner I, Mottaghy FM, Schreiber H, Riecker A, Ludolph AC, Kassubek J. Primary progressive aphasia accompanied by environmental sound agnosia: A neuropsychological, MRI and PET study. *Psychiatry Research-Neuroimaging,* 146: 191-197, 2006.

Van Bezooijen R, Gooskens C. Identification of language varieties - The contribution of different linguistic levels. *Journal of Language and Social Psychology,* 18: 31-48, 1999.

Van Borsel J, Janssens L, Santens P. Foreign accent syndrome: an organic disorder? *Journal of Communication Disorders,* 38: 421-429, 2005.

van Dommelen WA. The contribution of speech rhythm and pitch to speaker recognition. *Language and Speech,* 30: 325-338, 1987.

Van Lancker D, Kreiman J. Voice discrimination and recognition are separate abilities. *Neuropsychologia,* 25: 829-834, 1987.

Van Lancker DR, Canter GJ. Impairment of voice and face recognition in patients with hemispheric damage. *Brain and Cognition,* 1: 185-195, 1982.

Van Lancker DR, Kreiman J, Cummings J. Familiar voice recognition: patterns and parameters . *Journal of Phonetics,* 13: 19-38, 1985.

Van Lancker DR, Cummings JL, Kreiman J, Dobkin BH. Phonagnosia: a dissociation between familiar and unfamiliar voices. *Cortex,* 24: 195-209, 1988.

Van Lancker DR, Kreiman J, Cummings J. Voice perception deficits: neuroanatomical correlates of phonagnosia. *Journal of Clinical and Experimental Neuropsychology,* 11: 665-674, 1989.

Vandenberghe R, Price C, Wise R, Josephs O, Frackowiak RS. Functional anatomy of a common semantic system for words and pictures. *Nature,* 383: 254-256, 1996.

Vanstone AD, Cuddy LL. Musical Memory in Alzheimer Disease. *Aging Neuropsychology and Cognition,* 17: 108-128, 2010.

Vestergaard MD, Haden GP, Shtyrov Y, Patterson RD, Pulvermuller F, Denham SL, Sziller I *et al*. Auditory size-deviant detection in adults and newborn infants. *Biological Psychology,* 2009.

von Kriegstein K, Eger E, Kleinschmidt A, Giraud AL. Modulation of neural responses to speech by directing attention to voices or verbal content. *Brain Research Cognitive Brain Research,* 17: 48-55, 2003.

von Kriegstein K, Kleinschmidt A, Sterzer P, Giraud AL. Interaction of face and voice areas during speaker recognition. *Journal of Cognitive Neuroscience,* 17: 367-376, 2005.

von Kriegstein K, Kleinschmidt A, Giraud AL. Voice recognition and cross-modal responses to familiar speakers' voices in prosopagnosia. *Cerebral Cortex,* 16: 1314-1322, 2006.

von Kriegstein K, Warren JD, Ives DT, Patterson RD, Griffiths TD. Processing the acoustic effect of size in speech sounds. *Neuroimage,* 32: 368-375, 2006.

von Kriegstein K, Smith DRR, Patterson RD, Ives DT, Griffiths TD. Neural representation of auditory size in the human voice and in sounds from other resonant sources. *Current Biology,* 17: 1123-1128, 2007.

von Kriegstein K, Smith DR, Patterson RD, Kiebel SJ, Griffiths TD. How the human brain recognizes speech in the context of changing speakers. *Journal of Neuroscience,* 30: 629-638, 2010.

von Kriegstein KV, Giraud AL. Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *Neuroimage,* 22: 948-955, 2004.

Walden BE, Montgomery AA, Gibeily GJ, Prosek RA, Schwartz DM. Correlates of Psychological Dimensions in Talker Similarity. *Journal of Speech and Hearing Research,* 21: 265-275, 1978.

Waltz JA, Knowlton BJ, Holyoak KJ, Boone KB, Mishkin FS, Santos MD, Thomas CR *et al*. A system for relational reasoning in human prefrontal cortex. *Psychological Science,* 10: 119-125, 1999.

Wang XQ. On cortical coding of vocal communication sounds in primates. *Proceedings of the National Academy of Sciences of the United States of America,* 97: 11843-11849, 2000.

Warren JD, Griffiths TD. Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. *Journal of Neuroscience,* 23: 5799-5804, 2003.

Warren JD, Uppenkamp S, Patterson RD, Griffiths TD. Separating pitch chroma and pitch height in the human brain. *Proceedings of the National Academy of Sciences of the United States of America,* 100: 10038-10042, 2003.

Warren JD, Jennings AR, Griffiths TD. Analysis of the spectral envelope of sounds by the human brain. *Neuroimage,* 24: 1052-1057, 2005.

Warren JD, Scott SK, Price CJ, Griffiths TD. Human brain mechanisms for the early analysis of voices. *Neuroimage,* 31: 1389-1397, 2006.

Warren JE, Wise RJ, Warren JD. Sounds do-able: auditory-motor transformations and the posterior temporal plane. *Trends in Neurosciences,* 28: 636-643, 2005.

Warrington EK, James M. An experimental study of facial recognition in patients with unilateral cerebral lesions. *Cortex,* 3: 317-326, 1967.

Warrington EK. The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology,* 27: 635-657, 1975.

Warrington EK. Neuropsychological evidence for multiple memory systems. *Ciba Foundation symposium,* 69: 153-166, 1979.

Warrington EK. Recognition Memory Test. Windsor: NFER-Nelson; 1984.

Warrington EK, James M. Visual object recognition in patients with right-hemisphere lesions: axes or features? *Perception,* 15: 355-366, 1986.

Warrington EK, James M. Visual apperceptive agnosia: a clinico-anatomical study of three cases. *Cortex,* 24: 13-32, 1988.

Warrington EK, James M. The visual object and space perception battery. Bury St Edmunds: Thames Valley Test Company; 1991.

Warrington EK. The Graded Naming Test: a restandardisation. *Neuropsychological Rehabilitation,* 7: 143-146, 1997.

Warrington EK, McKenna P, Orpwood L. Single word comprehension: A concrete and abstract word synonym test. *Neuropsychological Rehabilitation,* 8: 143-154, 1998.

Watanabe T, Yagishita S, Kikyo H. Memory of music: roles of right hippocampus and left inferior frontal gyrus. *Neuroimage,* 39: 483-491, 2008.

Wechsler D. Wechsler Memory Scale-Revised. San Antonio: The Psychological Corporation; 1987.

Wechsler D. WAIS-III administration and scoring manual. San Antonio: The Psychological Corporation; 1997.

Wechsler D. Wechsler Abbreviated Scale of Intelligence. San Antonio: The Psychological Corporation; 1999.

Wessinger CM, Buonocore MH, Kussmaul CL, Mangun GR. Tonotopy in human auditory cortex examined with functional magnetic resonance imaging. *Human Brain Mapping,* 5: 18-25, 1997.

Whiteley AM, Warrington EK. Selective impairment of topographical memory: a single case study. *Journal of Neurology, Neurosurgery and Psychiatry,* 41: 575-578, 1978.

Whitwell JL, Crum WR, Watt HC, Fox NC. Normalization of cerebral volumes by use of intracranial volume: implications for longitudinal quantitative MR imaging. *American Journal of Neuroradiology,* 22: 1483-1489, 2001.

Whitwell JL, Przybelski SA, Weigand SD, Ivnik RJ, Vemuri P, Gunter JL, Senjem ML *et al*. Distinct anatomical subtypes of the behavioural variant of frontotemporal dementia: a cluster analysis study. *Brain,* 132: 2932-2946, 2009.

Wicklund AH, Johnson N, Rademaker A, Weitner BB, Weintraub S. Profiles of decline in activities of daily living in non-Alzheimer dementia. *Alzheimer Disease & Associated Disorders*, 21: 8-13, 2007.

Wildgruber D, Pihan H, Ackermann H, Erb M, Grodd W. Dynamic brain activation during processing of emotional intonation: influence of acoustic parameters, emotional valence, and sex. *Neuroimage,* 15: 856-869, 2002.

Wildgruber D, Riecker A, Hertrich I, Erb M, Grodd W, Ethofer T, Ackermann H. Identification of emotional intonation evaluated by fMRI. *Neuroimage,* 24: 1233-1241, 2005.

Wildgruber D, Ackermann H, Kreifelts B, Ethofer T. Cerebral processing of linguistic and emotional prosody: fMRI studies. *Progress in Brain Research,* 156: 249-268, 2006.

Williams GB, Nestor PJ, Hodges JR. Neural correlates of semantic and behavioural deficits in frontotemporal dementia. *Neuroimage,* 24: 1042-1051, 2005.

Wilson SM, Dronkers NF, Ogar JM, Jang J, Growdon ME, Agosta F, Henry ML *et al*. Neural correlates of syntactic processing in the nonfluent variant of primary progressive aphasia. *Journal of Neuroscience,* 30: 16845-16854, 2010.

Wong PC, Parsons LM, Martinez M, Diehl RL. The role of the insular cortex in pitch pattern perception: the effect of linguistic contexts. *Journal of Neuroscience,* 24: 9153-9160, 2004.

Wu K, Childers DG. Gender recognition from speech. Part I: Coarse analysis. *Journal of the Acoustical Society of America,* 90: 1828-1840, 1991.

Young AW, McWeeny KH, Hay DC, Ellis AW. Access to identity-specific semantic codes from familiar faces. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology,* 38: 271-295, 1986.

Young AW, Newcombe F, de Haan EH, Small M, Hay DC. Face perception after brain injury. Selective impairments affecting identity and expression. *Brain,* 116: 941-959, 1993.

Zatorre RJ, Evans AC, Meyer E, Gjedde A. Lateralization of phonetic and pitch discrimination in speech processing. *Science,* 256: 846-849, 1992.

Zhang LJ, Shu H, Zhou FY, Wang XY, Li P. Common and distinct neural substrates for the perception of speech rhythm and intonation. *Human Brain Mapping,* 31: 1106-1116, 2010.

Zhou J, Greicius MD, Gennatas ED, Growdon ME, Jang JY, Rabinovici GD, Kramer JH *et al*. Divergent network connectivity changes in behavioural variant frontotemporal dementia and Alzheimer's disease. *Brain,* 133: 1352-1367, 2010.