AUDITORY COMPREHENSION: FROM THE VOICE UP TO THE SINGLE WORD LEVEL

Volume 1

Anna Barbara Jones

THE UNIVERSITY
of EDINBURGH

Doctor of Philosophy

The University of Edinburgh

2016

Declaration

I declare that this thesis was composed by me and that the work contained therein is my own, expect where explicitly stated otherwise in the text. I further declare that the work within this thesis has not been submitted for any other degree or professional qualification.

(Anna Barbara Jones)

6th March 2016

Abstract

Auditory comprehension, the ability to understand spoken language, consists of a number of different auditory processing skills. In the five studies presented in this thesis I investigated both intact and impaired auditory comprehension at different levels: voice versus phoneme perception, as well as single word auditory comprehension in terms of phonemic and semantic content.

In the first study, using sounds from different continua of ‘male’-/pæ/ to ‘female’-/tæ/ and ‘male’-/tæ/ to ‘female’-/pæ/, healthy participants (n=18) showed that phonemes are categorised faster than voice, in contradistinction with the common hypothesis that voice information is stripped away (or normalised) to access phonemic content. Furthermore, reverse correlation analysis suggests that gender and phoneme are processed on the basis of different perceptual representations. A follow-up study (same paradigm) in stroke patients (n=25, right or left hemispheric brain lesions, both with and without aphasia) showed that lesions of the right frontal cortex (likely ventral inferior frontal gyrus) leads to systematic voice perception deficits while left hemispheric lesions can elicit both voice and phoneme deficits. Together these results show that phoneme processing is lateralized while voice information processing requires both hemispheres. Furthermore, this suggests that commencing Speech and Language Therapy at a low level of acoustic processing/voice perception may be an appropriate method in the treatment of phoneme perception impairments.

A longitudinal case study (CF) of crossed aphasia (rare acquired communication impairment secondary to lesion ipsilateral to the dominant hand) is then presented alongside a mini-review of the literature. Extensive clinical investigation showed that CF presented with word-finding difficulties related to impaired auditory phonological analysis, while functional Magnetic Resonance Imaging (fMRI) analyses showed right hemispheric lateralization of language functions (reading, repetition and verb generation). These results, together with the co-morbidity analysis from the mini-review, suggest that crossed aphasia can be explained by developmental disorders which cause partial right lateralization shift of language processes. Interestingly, in CF this process did not affect voice lateralization and information processing, suggesting partial segregation of voice and speech processing.

In the last two studies, auditory comprehension was examined at the single word level using a word-picture matching task with congruent (correct target) and incongruent (semantic, phonological and unrelated foils) conditions. fMRI in healthy participants (n=16) revealed a key role of the pars triangularis (phonological processing), the left angular gyrus (semantic incongruency) and the left precuneus (semantic relatedness) in this task – regions typically associated via the arcuate fasciculus and often impaired in aphasia. Further investigation of stroke patients on the same task (n=15) suggested that the connections between the angular gyrus and the pars triangularis serve a fundamental role in semantic processing. The quality of a published word-picture matching task was also investigated, with results questioning the clinical relevance of this task as an assessment tool.

Finally, a pilot study looking at the effect of a computer-assisted auditory comprehension therapy (React2©) in 6 stroke patients (vs. 6 healthy controls and 6 stroke patients without therapy) is presented. Results show that the more therapy patients carry out the more improvement is seen in the semantic processing of single nouns. However, these results need to be reproduced on a larger scale in order to generalise any outcomes.

Overall, the findings from these studies present new insight into, as well as extending on, current cognitive and neuroanatomical models of voice perception, speech perception and single word auditory comprehension. A combinatorial approach to cognitive and neuroanatomical models is proposed in order to further research, and thus improve clinical care, into impaired auditory comprehension.

Lay Summary

Auditory comprehension is the ability to understand spoken language, and is a skill consisting of a number of different processes including a) the perception of voices, b) perception of speech sounds and c) understanding of spoken single words in terms of semantic (the meaning of words) and phonemic (speech sounds) content. The aim of the 5 different studies presented within this thesis is to investigate auditory comprehension in the healthy population as well as in patients with brain damage as a result of a stroke, at these three levels.

The first study in this thesis looks at auditory comprehension at level of the voice (skill a: by using male/ female gender) and speech (skill b: by using the speech sounds 'pa' and 'ta') in the healthy population. The results suggest that our brains do not firstly strip away gender content in order to process speech sound content (as suggested in previous studies). Further analysis suggests that when we are trying to decide if the voice we hear is male or female (gender categorisation) our brains use different cues to when we are trying to process speech sounds. The second study uses the same task but in stroke patients, with right or left hemisphere brain lesions, both with and without aphasia (an acquired communication impairment as a result of a stroke). Note that patients with right hemisphere damage and aphasia were not recruited because this is a very rare impairment called crossed aphasia – discussed in the following case study. Results show that we use both hemispheres of our brains when we process voices, and also that the right frontal cortex of the brain is critical in carrying out this process. This also suggests that in clinical practice, Speech and Language Therapists may be better focussing on therapy that helps stroke patients to process voices, rather than jumping ahead to process language in its entirety.

Thirdly, a single case study is presented looking at an individual (CF) with a crossed aphasia (a type of aphasia that occurs secondary to a lesion in the right hemisphere in a right-handed individual), alongside a mini-review of the literature. Extensive clinical investigation showed that CF presented with word-finding difficulties, while functional Magnetic Resonance Imaging (fMRI) analyses showed that certain language functions (reading, repetition and verb generation) were lateralized in the right hemisphere of his brain. Together the results of the literature review and the case study investigations suggest that crossed aphasia can be caused by a developmental disorder which in turn causes a shift of language processes in the brain from the left to right hemisphere. Interestingly, in CF this shifting process did not affect his voice perception, suggesting partial separation of voice and speech processing.

In the last two studies, auditory comprehension is looked at in terms of the processing of single words, in healthy participants and in stroke patients. Participants were asked to carry out a word/picture matching task when inside an fMRI scanner. The fMRI in healthy participants revealed that key areas of the brain are involved in this task, and that these regions are often impaired in aphasia. Results from stroke patients (same task) suggest that particular connections between brain regions are vital to understand single word semantics.

Finally, a pilot study looking at the effect of an additional speech and language therapy given to patients with aphasia (and compared to both healthy controls and patients with aphasia not given the therapy) on a computer is presented. Results show that the more therapy patients carry out the more improvement is seen in their ability to understand single word semantics. However, these results need to be reproduced on a larger scale in order to be able to generalise any outcomes.

Overall, this thesis presents studies and results that further our understanding of auditory comprehension, as well as providing some suggestions that may be helpful to speech and language therapists when assessing and treating patients who have communication impairments.

Acknowledgments

This PhD thesis would not have been possible without the help of a great number of people, in so many ways.

Firstly, I would like to thank my primary supervisor, Dr. Cyril Pernet, without whom this thesis would not have been possible. Thank you for all your help, support, guidance, time, vast knowledge and patience over the years. Thank you also to my second supervisor, Dr. Thomas Bak, for all your enthusiasm, support and motivation throughout my PhD.

I would like to thank the Brain Research Imaging Centre, University of Edinburgh and the Division of Clinical Neurosciences, Western General Hospital for allowing me to undertake my work here. Many thanks also go to the SINAPSE Collaboration (a Pooling Initiative funded by the Scottish Funding Council and the Chief Scientist Office of the Scottish Executive) and the Tony Watson Scholarship for providing the financial support for my PhD.

I am extremely grateful to all the participants who took part in these studies, attending many appointments and investigations, often at a very worrying and difficult time – thank you.

I would like to thank all my colleagues who have been with me on this journey over the years, helping me in a huge variety of ways. Thank you especially to Moira Henderson, Dr. Duncan Martin, Dr. Mark Bastin, Dr. Dominic Job, Prof. Andrew Farrall and Prof. Joanna Wardlaw. Special thanks also go to all the radiographers (Elaine Sanderman, Gail Barclay, Iona Hamilton, Charlotte Jardine and Stewart Wiseman) who made it possible for patients to be scanned, fed me many a cake and kept me smiling throughout this process.

Thanks also go to the Speech and Language Therapists within NHS Lothian who willingly provided their expert knowledge, helped me with the recruitment of my participants, and generously allowed me to borrow rooms and resources.

I would like to thank all my close lab mates (Anna Heye, Dr. David Dickie, Dr. Andreas Glatz, Dr. Natalie Royle, Dr. Jehill Parikh and Dr. Xin Wang) for keeping me sane, not letting me quit, and becoming life-long friends.

A big thanks also go to my friends outside of the PhD world, who endured many years of moaning and spent endless days encouraging me and helping me to forget about it all for a while.

I would like to thank my family for all their love and support throughout this process– thanks especially go to my amazing mum, my wonderful dad, my inspiring brother (Owen), my two supportive sisters (Laura and Sonia) and my brothers-in-law (Gordon and Will), as well as my in-laws on the other side (Elaine, Byron, Phil, Louise, Ali, Erica, Fraser and Eilidh) who have stood with me and supported me through the good times and the bad.

Finally, I would like to thank my husband, Peter McKibben. Thank you for helping me immeasurably over the years – I literally could not have done this without you and your unfailing faith, love, trust and support. Thank you for standing by me always.

This thesis is dedicated to Iona
"too beautiful for earth"

Table of contents

List of figures

List of tables

1. Chapter 1 – Introduction to Auditory Comprehension

1.1. Introduction to auditory processing

The focus of this thesis is intact and impaired auditory comprehension: the ability to understand spoken language. Auditory comprehension is not, however, a standalone process – there are a number of auditory processing skills, as well as other cognitive counterparts, underpinning comprehension. In the initial studies presented here, I looked at auditory perception of voices compared to auditory perception of phonemes and whether impairments in these areas correspond to impaired auditory comprehension skills. The studies investigated this in both healthy controls and stroke patients, with right or left hemisphere brain lesions, both with and without aphasia (acquired communication impairment). This is an important and often neglected area of research. The majority of studies investigate comprehension deficits without asking if more primary impairments (i.e. voice perception deficits) can explain the pattern of performances. A longitudinal single case study is also presented inquiring voice perception, speech perception and auditory comprehension in an individual with a crossed aphasia (rare acquired communication impairment secondary to a lesion in the right hemisphere in a right-handed individual). This offers a different perspective on the relationships between structures and functions and the role of plasticity. Finally, MRI- and fMRI-based studies looking at the neural substrates of auditory comprehension in healthy participants and left hemispheric stroke patients with aphasia are presented along with behavioural results from a pilot study on computer-assisted therapy in a subgroup of these patients.

In this 1st chapter, I present some background information on auditory comprehension and processes underpinning it.

1.2. The auditory system

The auditory system involves a number of functionally distinct peripheral and central components. In humans the ability to perceive sound relies on the detection of pressure changes causing vibrations in the ear. The external ear and ear canal entrance are the only parts of the auditory system visible on the outside of the body. The external ear's role is to gather sound waves and direct them into the ear canal. These waves then hit the tympanic membrane (eardrum). The subsequent vibrations then activate the ossicular chain: malleus, incus and stapes (Figure 1-1). These are three tiny, connected bone structures that form the middle ear. The stapes vibrations are then transmitted to the cochlear fluids through the oval window. At this point, sound wave information enters the inner ear, and changes from mechanical to neuroelectric (Warren et al., 2009).

The inner ear contains the cochlea, basilar membrane, auditory receptor cells and the auditory nerve. The cochlear is a small (~10mm wide) but complex coiled structure containing the organ of Corti (hearing organ), vestibule (part of the balance-regulating system), oval and round windows (membrane-covered pressure-regulation system), and the semi-circular canals (part of the balance-regulating system) (Warren et al., 2009).

When vibrations reach the inner ear they are propagated through the cochlear fluid-membrane system, containing the endolymph and perilymph. This physical force causes the basilar membrane to vibrate in a tonotopic pattern (Figure 1-1) – each frequency has a place of resonance along the membrane, with high frequencies resonating at the end near the basal entrance to the cochlea, and low frequencies at the apex (Boer, 1976; Talavage et al., 2004). The movement of the basilar membrane moves the attached hair cells (stereocilia) against the tectorial membrane. This causes the hair cells to become depolarised, leading to afferent signals in the cochlear nerve.

The central auditory pathway contains complex connections transferring information from the organ of corti to the primary auditory cortex in the temporal lobe of the brain. Information is first passed up to the dorsal and ventral cochlear nuclei in the brainstem, where the afferent fibres of the vestibulocochlear nerve (VIIIth cranial nerve) terminate. The majority of fibres then pass to the superior olivary nucleus in the pons, and to the inferior colliculus in the midbrain. Information then leaves the brainstem to project onto the medial geniculate body in thalamus to finally project onto the primary auditory cortex (Figure 1-2).

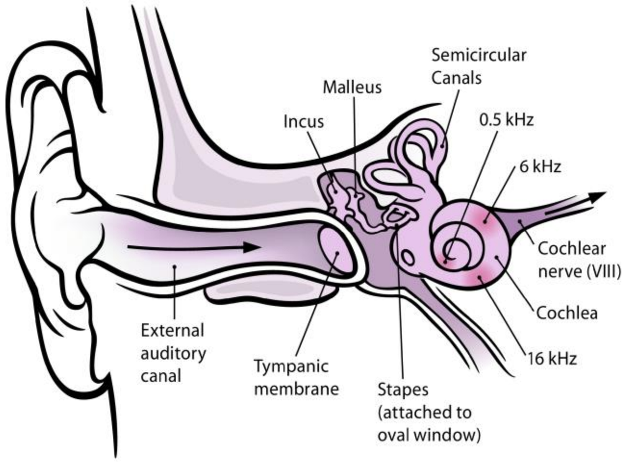

Figure 1-1: The human ear and frequency mapping in the cochlear.
Taken from Chittka & Brockmann (2005)

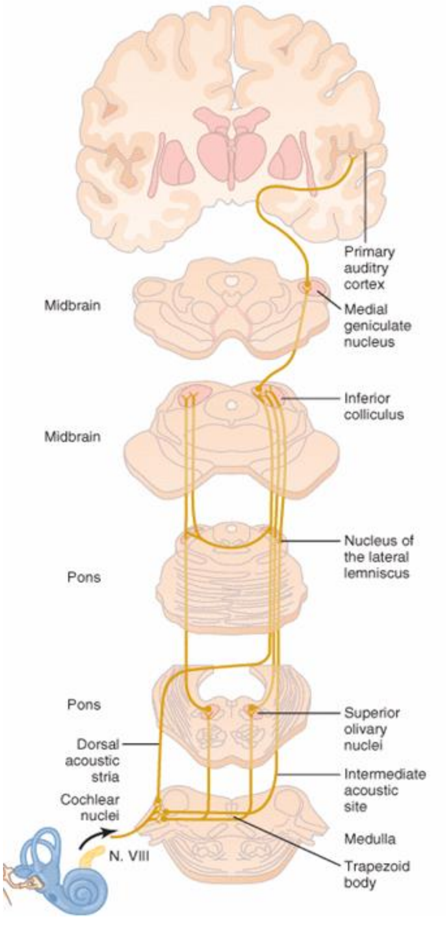

Figure 1-2: Central Auditory Pathways.
The central auditory pathways extending from the cochlear nucleus to the primary auditory cortex. Taken from: Hall and Guyton (2005)

1.3. Perception of speech: current approaches

Speech perception involves audition of the sound signal, followed by the extraction of acoustic cues to gain phonetic and linguistic information. This information is then used for higher-level language processes, for example word recognition. Although much research has been done in the area of speech perception, it is still not understood fully, mainly because speech is highly variable (due to different phonetic environments, different speech conditions and individual speaker identities). The first hypothesis of the neurology of speech perception was suggested over 100 years ago by Carl Wernicke (1874); complete consensus has yet to be reached.

Current approaches and theories of speech perception are consistent in their involvement of the mapping of speech acoustic signals onto linguistic messages, i.e. phonemes, syllables, words etc. However, the way in which this occurs has been debated for centuries, and there remain a number of plausible approaches and theories.

Exemplar models of speech perception (also referred to as episodic theories) build upon the hypothesis that listeners store speech sounds in their memory. When perceiving speech, the listener activates the unique stored memory trace containing the speech information and compares the incoming stimulus to stored representations. Exemplars are the storage of words and high-frequency grammatical constructions within the memory (Bybee and Hopper, 2001; Coleman, 2002). They are highly context-dependent (Docherty and Foulkes, 2004). Bradlow et al. (1999) showed, for instance, that listeners are more accurate at recognition of a repeated word when the stimulus is from the same speaker at the same output rate, therefore confirming the idea of the retention of a surface form. This theory has been supported by studies showing that an unfamiliar talker, unpredictable talker output, or misidentification of a speaker's sex leads to a higher error rate in word identification (Lightfoot, 1989; Strand, 2000; Walker et al., 1995). Exemplar theories can also account for top-down effects such as the Ganong effect (Ganong, 1980) and, if extended to a multimodal exemplar theory, the McGurk effect (McGurk and MacDonald, 1976). There are three main assumptions that need to be taken into account when discussing exemplar models. First, that the patterns of lexical activation are influenced by fine phonetic detail. Secondly, that the properties of the auditory stimuli and stores are context-dependent. Third, that abstract entities cannot be assigned and therefore stored (Nguyen et al., 2009). Aside from these assumptions there are a number of objections to models based on exemplar theory (Lahiri and Reetz, 2002; Nguyen et al., 2009; Nygaard et al., 1994; Pallier et al., 2001; Peperkamp, 2003). A major argument against the exemplar theory is that humans would have insufficient memory capacity to store every token of sound they have ever heard (Johnson and Mullennix, 1997), especially considering the level of allophonic variation. This phenomenon accounts for speaker variability, whereby a listener interprets multiple allophones as the same phoneme, supporting the idea that the processing of phonemes becomes separate from the processing of acoustic details. Studies looking at identification and recognition of words and pictures (Goldinger, 1996) suggest, however, that humans have a large memory capacity and therefore large demands on memory capacity alone are not enough to disqualify the theory. Nonetheless, there are objections to connectionist models that involve storing each exemplar, on the grounds of memory saturation. However, the ALCOVE (attention learning covering map) connectionist exemplar model (Kruschke, 1992) proposes a perceptual space, or ‘covering map’, with locations on the map corresponding to vectors of possible auditory properties. Incoming auditory properties are connected to locations of possible exemplars on this bounded map, with each location on the map connected to category nodes. This allows for exemplars to be coded rather than stored explicitly, thus negating the possible problem of memory saturation (Johnson, 1997).

Abstract models of speech perception focus on abstract phonological units or lexical forms, which represent information needed to differentiate between citation-form words (i.e., the lexical set of a word form). This approach states that early abstractions are ordered into formal linguistic units with the features/phonemes preserved and the variation discarded – such models were developed to account for the aforementioned allophonic variations. The Motor theory (Liberman et al., 1967) is an example of a theory based upon the abstractionist approach. Over the past two decades, episodic/exemplar and abstractionist approaches to speech perception have been strongly debated, with many studies suggesting that different approaches support differently the various properties of speech perception. The traditional abstractionist view of speech perception, which relies on sequences of discrete units, has had a long-lasting effect on the field of speech perception. With advances in cognitive neuropsychology, however, it has become clearer that the episodic approach is also very valid within this debate. Goldinger (2007) brings a complementary approach to this debate, suggesting that an optimal theory may include both exemplars, in the form of episodic traces, and stable abstract representations. This paper also supports a complementary neural network and anatomical approach to speech perception. McQueen et al. (2006) also suggest that "hybrid abstractionist-episodic models, therefore hold considerable promise", a view also held by a number of others (Foulkes and Docherty, 2006; Nguyen et al., 2009; Smith, 2004).

1.4. Key theories of speech perception

Three major theories dominate research:

- Motor theory (abstractionist model)
- Direct realism (exemplar model)
- General approach/auditory theory (exemplar-abstractionist model)

Liberman & Mattingly (1985) formally proposed the **motor theory** of speech perception, following investigations and hypotheses by Liberman & Cooper in the 1950s. Motor theory hypothesizes that humans perceive spoken words through the identification of the intended gestures of the vocal tract producing the speech, rather than through the identification of the sound patterns produced by the spoken output. The discovery of mirror neurons - neurons that link perception of motor movements with production of motor movements - has further advanced the motor theory of speech. These neurons were first discovered in the monkey premotor cortex and are activated both when monkeys perform an action, and when they observe another individual carrying out a similar action (di Pellegrino et al., 1992; Gallese et al., 1996; Rizzolatti et al., 1996). Rizzolatti & Craighero (2004) provided a review of mirror neuron properties within the human cortex, linking the mirror-neuron system to language by building on a data set published by Rizzolatti (1998). These papers show that there is an ‘echo-neuron system’ that causes speech-related motor activation centers to be activated when humans listen to verbal stimuli (Rizzolatti and Craighero, 2004). The core of the neuroanatomy of the mirror-neuron system within humans was proposed to include major sections of the frontal lobe, in particular the posterior part of the inferior frontal gyrus (IFG), the lower part of the precentral gyrus (PCG), as well as the anterior section of the inferior parietal lobe (Rizzolatti and Craighero, 2004). In terms of specific speech related areas of this system, Buccino (2001) suggested that all these areas are activated bilaterally, findings that are supported by Pazzaglia (2008) in a voxel-based lesion symptom mapping (VLSM) study. However, the proposed neural substrates of the mirror-neuron system have attracted opposition. Molenberghs (2012) reviews numerous studies that together suggest the involvement of 34 different Brodmann areas, implying an extensive network of activity in this system. Because the likelihood of all these areas possessing mirror neurons is limited, a refined core network of the following brain areas was proposed by Molenberghs et al. (2012): IFG, dorsal and ventral PCG, the inferior and superior parietal lobule, and a number of additional areas specific to the sensory system involved (auditory, somatosensory, or emotional processing).

The **direct realism** theory of speech perception is derived from Gibson’s (1966) perceptual theory and was proposed by Fowler in the 1980s (Fowler, 1986; Fowler and Dekle, 1991). It suggests that the objects of perception are not the abstract phonemes or intended events (as with the motor theory), but rather the actual vocal tract movements. Therefore, listeners perceive these gestures via the information provided in the acoustic signal that specifies the gestures that form this signal. Thus, this theory does not involve a specialized system of decoding abstract phonemes, as with the motor theory. As for the motor theory of speech, mirror neurons are viewed as providing a mechanism for linking perception and action, not solely in speech but in perception generally (Galantucci et al., 2006).

Fowler & Rosenblum (1990) demonstrated their direct realism standpoint through a study examining duplex perception. Duplex perception is the simultaneous perception of both speech and non-speech stimuli extracted from segments of the speech signal e.g. /da/ versus /ga/ which differ in the transition of the third formant. However, the remainder of the syllable is the same in both instances, and is perceived as an ambiguous syllable. On their own, the third formant transitions are perceived as ‘chirps’. Dichotic listening tasks showed that simultaneous presentation of a ‘chirp’ and the ambiguous remainder syllable were perceived as either /da/ or /ga/, depending on the onset frequency of the third formant transition. Fowler & Rosenblum (1990) interpret this to indicate that meaningful sounds (i.e., speech) take precedence over non-meaningful sounds (i.e., chirp). Further support for the direct realism theory comes from McGurk & MacDonald (1976), who demonstrated that speech perception uses a compromise of visual and auditory inputs (the McGurk effect).

The **general auditory approach** to speech perception (Diehl et al., 2004) simply suggests that speech sounds are perceived via the same mechanisms as any other sounds. Thus, it does not include gestures (speech or intended) within perception and no specialized modules or mechanisms are used. This approach is not a theory, but rather a general framework proposed as an alternative to both the motor and the direct realism theories of the late 1970s, and support for it has grown ever since (Diehl, 1987; Diehl and Kluender, 1989; Holt et al., 1998; Kingston and Diehl, 1995; Kluender and Lotto, 1994; Lotto, 2000; Massaro and Oden, 1980; Miller and Johnson-Laird, 1976; Pisoni, 1977; Stevens and Klatt, 1974; Warren et al., 2009). Stevens and Klatt (1974) and Pisoni (1977) found that both speech and non-speech stimuli share temporal properties that are essential to perception, and suggested that general auditory mechanisms are responsible. Kuhl & Miller (1978) looked at speech perception in non-human animals, and found that similar aspects are exhibited, supporting the general auditory approach. In Kluender’s (1987) study into speech perception in birds, it was found that these animals could be trained to respond to varying syllables with initial consonant /d/ and not to those with initial consonant /b/ or /g/, despite the fact that such animals have no speech production. This provides supportive evidence that articulatory gestures and specialized speech perception modules are not required for perception even when there is acoustic variability. This phenomenon is also found in humans who are able to perceive syllable initial constants as equal despite acoustic variance.

1.5. Categorical perception of speech perception

When listening to somebody speaking, sounds are perceived as linguistic units within an extreme range of acoustic variation. Categorical perception refers to this ability to assign functionally equivalent classes to potentially discriminate speech sounds. It has been proposed that general auditory mechanisms contribute to the categorical perception of speech sounds: (i.e. cross-linguistic VOT studies -Eimas et al., 1971; infant VOT studies -Kuhl and Padden, 1983; phoneme discrimination in macaques -Lisker and Abramson, 1963).

In this thesis, the definition of categorical perception is based on continua of stimuli, varying along one or more parameters. These continua are divided by a perceptual boundary, such that stimuli on one side of the boundary are perceived as belonging to one category and stimuli on the other side are perceived as belonging to another category, despite physical linear spacing. Moreover, discrimination between stimuli from the same category must be low compared to discrimination across the boundary; if there is no discrimination peak at the boundary, perception is continuous rather than categorical.

A large amount of studies looked at the categorical perception of voiced versus voiceless initial stop consonants with varying voice-onset time. Voice-onset time is the time between the end of the stop-consonant noise burst and the beginning of periodicity. Studies on temporal order detection showed that an onset asynchrony of 20 milliseconds (msec) is needed. Thus, the minimum detectable VOT's must be +/20 msec. This mechanism, however, is essentially an acoustic parameter that depends on threshold mechanisms. In addition, it is not specific to humans; Kuhl & Miller (1978) reported that it is possible to train chinchillas in such a way that they achieve human performance in identifying stimuli along a VOT continuum. Another area of investigation focuses on the place of articulation in the vocal tract of stop consonants with varying second- and third-formant frequency transitions. In order to go from a particular stop consonant to a ‘steady-state’ vowel, or from a vowel to a final stop consonant, the speech mechanism has to perform particular movements. These movements result in consonant specific formant transitions, therefore the normal place-of-articulation continuum can be divided into three categories: /p/, /t/, /k/ for voiceless consonants and /b/, /d/, /g/ for voiced ones. Results from categorical experiments are often used to support the motor theory or direct realism theory of language. This is because, as there is no articulatory continuum from, for example, /p/ via /t/ to /k/, therefore when listening in speech mode we are compelled to divide the acoustic continuum into three articulatory categories. However, rapid transients in complex stimuli can be perceived categorically, whether they are speech or non-speech. Cutting & Rosner (1974) showed that by varying rise times, a speech continuum could be created, ranging from /tʃ/ to /ʃ/ (e.g. from chip to ship: very short rise times are perceived as /t/ before /ʃ/); a sawtooth waveform treated in the same way gave rise to responses ranging from ‘plucked’ to ‘bowed’ (violin tone). It seems, therefore, that phoneme categorization is based on thesholding mechanisms allowing boundaries to form between acoustic cues. This approach is challenged, however, by the large degree of overlap between cue distributions from various speakers (Holt and Lotto, 2010).

Other higher-level effects have contributed to the debate surrounding speech perception and speech categorization theories, and also challenge ‘simple’ boundary conceptualizations. These include the commonly known McGurk effect (1976); the Ganong effect (Ganong, 1980); and phonemic restoration (Warren, 1970). These effects demonstrate that speech perception cannot be understood solely in terms of the acoustic or gestural event and the perception of the phonetic content; it also relies on additional sources of information, and several levels of processing.

A number of subsequent models have been proposed (detailed below), building on the initial/more generalized speech perception theories proposed earlier, incorporating other relevant sources of information.

1.6. Models of spoken word recognition: advances on the traditional theories

The Cohort model is a lexical retrieval theory first proposed by Marslen-Wilson (1987), which looks at the mapping of the auditory (and visual) input onto a word in the lexicon of the listener. The model proposes that speech segments activate all words in the lexicon that commence with the initial section of that particular speech segment. As more segments are added, more words in the lexicon are thus 'ruled out', until only one word is left that matches the incoming speech signal. There are three main sections to the cohort model, all of which are bottom-up processing: first, the 'access stage' when the lexicon activates all possibly words; second, the 'selection stage' when activation decreases on word no longer selected due to additions in the speech segment; and third, the 'recognition point or 'uniqueness point' when only one word remains activated and is thus selected. The model has been adjusted to include a further 'integration' stage, where semantic and syntactic properties of the selected word are included in high-level representations of the word (Altmann, 1995). This top-down processing step allows for other features, including context, to play an integral role when ruling out competitors (Grosjean, 1980).

Connectionist models (i.e., neural network models) of speech perception suggest that interconnected networks of simple units carry out the processing (Bishop, 1995). The TRACE model is a popular connectionist model of speech perception. Proposed by McClelland & Elman (1986), the TRACE model builds upon the idea of the Ganong effect (Ganong, 1980). It is a neural network model that learns to examine patterns to classify the auditory input. Many words are activated simultaneously and the most highly activated word is identified as the most probable corresponding representation, and therefore partial information can activate a full lexical representation. There are many levels - phonetic features, phonemes, words - all of which are fixed. The Adaptive Resonance Theory (Grossberg, 1980) is a very similar connectionist model, with one main difference: it assumes that the speech segments are chunked in the same way that a listener hears them. These are both interactive models using bottom-up and top-down information to carry out the processing.

Opposing models, using solely bottom-up processing, have been proposed by Norris and others (Norris, 1994; Norris et al., 2000; Norris et al., 2001). Such models include Shortlist and Merge models (Norris, 1994; Norris et al., 2000), and the Fuzzy Logical Model of Perception (Massaro, 1989). A schematic showing the difference between connectionist models (interactive top-down versus autonomous bottom-up) and cohort models (bottom-up processing) of speech perception is shown in Figure 1-3.

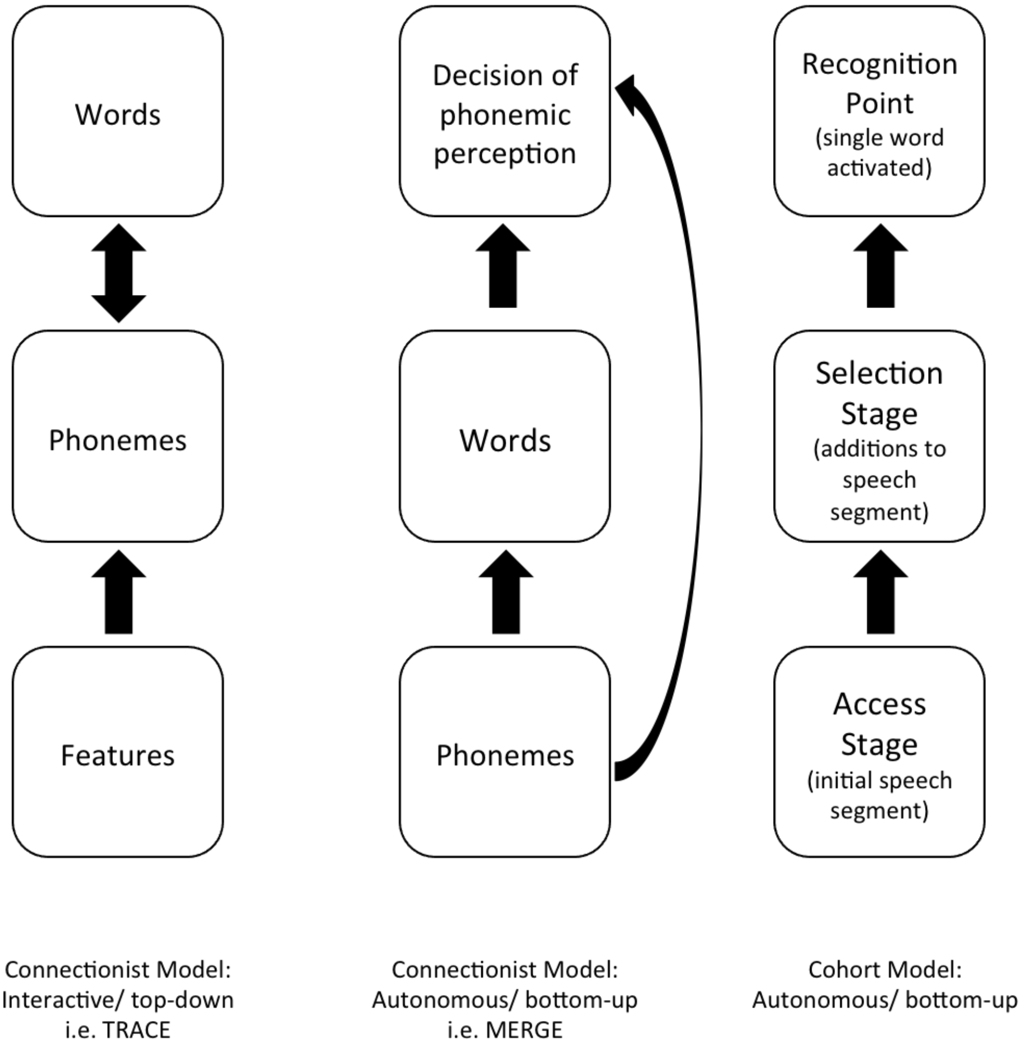

Figure 1-3: Schematic showing the difference between Connectionist and Cohort models (left – right) 1. Interactive connectionist models (top-down and bottom-up processing with feedback); 2. Autonomous connectionist models (bottom-up information only with no feedback); 3. Autonomous

cohort models (bottom-up processing) of speech perception. An example of an interactive connectionist model of speech perception is the TRACE model (McClelland and Elman, 1986); an example of an autonomous connectionist model of speech perception is the Merge model (Norris, 1999).

The Shortlist model is a popular connectionist model proposed by Norris (1994), containing lexical competition and solely bottom-up processing. It is a model of how people recognize words within continuous speech. This model was built upon shortly after its initial development, and Norris subsequently proposed the Merge model of speech perception (Norris, 1999). The Merge model contains two separate streams of perceptual analysis - phonemic code production and lexical output - and does not incorporate feedback (Figure 1-4). More recently, Norris & McQueen produced a new version of these models called Shortlist B which includes Bayesian inference, i.e., listeners use all the information available to them to derive optimal interpretations of the speech input unit (Norris and McQueen, 2008).

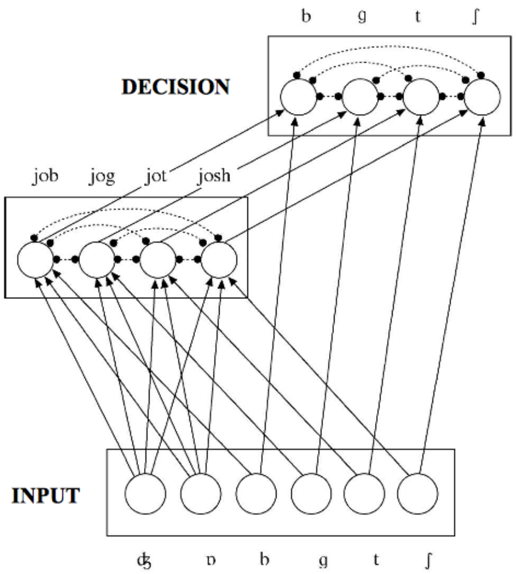

Figure 1-4: Schematic of the original Merge model of speech perception. Taken from Norris (1999).

The fuzzy-logical model was proposed by Massaro (1989), and suggests that categorical perception is not to be taken in the literal way, but as a way of classifying features in the speech pattern. Massaro (1989) proposes that individuals remember speech sounds as prototypes, and that speech perception occurs in three main stages:

- analysis of speech in terms of auditory and visual features
- integration of the information gained from the analysis
- identification of the speech sound using the information from the steps above.

The ‘fuzzy’ part of this model refers to the fact that the ‘units’ of the features are not straightforward true and false (i.e., 1s and 0s) but are ‘fuzzy’. This ‘fuzzy’ unit value is decided according to how likely it is that the sound belongs to a particular speech category. The units are therefore relative matches between the incoming features and the values of the prototypes.

1.7. Neuroanatomy of speech perception

The auditory system has been studied more extensively in macaques than in humans. These studies have provided information regarding the way the auditory cortex is organised (Benson et al., 2001; Liebenthal et al., 2010). In primates the auditory cortex consists of three main areas: ‘core’, ‘lateral belt’ and ‘parabelt’, each receiving input from the auditory thalamus (Kaas and Hackett, 2000; Rauschecker, 2011; Scott and Wise, 2003; Sekiyama et al., 2003). A subdivision within the ‘parabelt’ has been further identified (Sweet et al., 2005). The regions are distinguished according to their connections with each other, their connections with other auditory regions, and the source of their thalamic afferents (Hackett et al., 1998; Jones et al., 1995; Romanski et al., 1999).

Investigations into the auditory system in humans have been furthered through the use of MR imaging and, as a result, there is current evidence that these distinct areas of the auditory cortex also exist in humans (Meister et al., 2007). Bilaterally, Heschl’s gyrus (located superiorly within the temporal lobe) has been identified as the primary auditory cortex (PAC) (Scott and Johnsrude, 2003). The PAC in humans also has three areas located next to Heschl’s gyrus that are involved in this early stage of processing: the planum temporale; the planum polare; and the lateral STG (superior temporal gyrus - both anterior and posterior) extending to the STS (superior temporal sulcus)

(Morosan et al., 2001). Sweet et al. (2005) localised the 'core', 'lateral belt' and subdivisions of the 'parabelt' within the human STG, as shown in Figure 1-5.

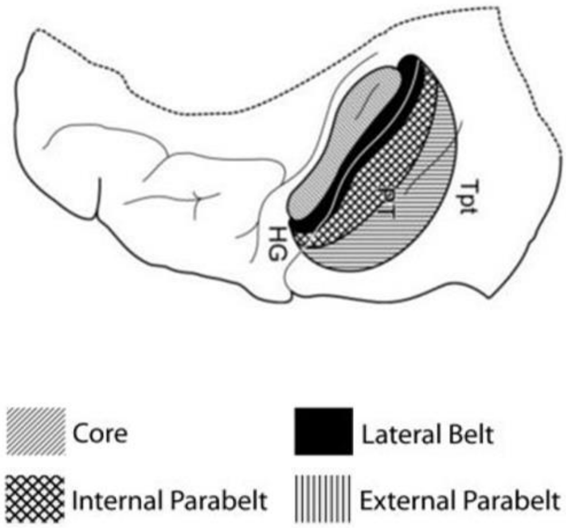

Figure 1-5: Diagram showing borders of the left auditory core, lateral belt and parabelt (internal and external subdivisions) in the human brain.
Rostral = left; caudal = right; lateral = bottom; medial = top. (HG: Heschl's gyrus, PT: Planum temporale). Image taken from Sweet et al. (2005).

There are a number of additional areas, both within and without the main auditory regions in the human brain (rather than the primate brain), that are involved in auditory perception and others that are involved in speech perception. As our structural understanding improves - and as the idea of a hierarchical structure has appeared - the 'dual-processing' models (with 'what' and 'where' streams) have become an established hypothesis of the auditory and speech pathways. It is helpful, therefore, to look at the human neuroanatomy of both auditory and speech perception through dual-processing models.

In the 1870s, Karl Wernicke proposed the first model of speech perception (Wernicke, 1874), stating that it is supported by the auditory cortex and, in particular, the involvement of the left STG in auditory comprehension. However, it was later found that large lesions to the left STG did not actually lead solely to auditory comprehension difficulties, but more commonly to speech production deficits (Damasio and Damasio, 1980). This led to ambiguity regarding the involvement of the STG in speech perception. Importantly, Wernicke's initial hypothesis included two streams of auditory information processing and became the basis of the dual-stream models of speech perception.

The **auditory dual-processing hypothesis** states that the auditory system can be thought of as two streams of processing, similar to those present in the visual and somatosensory systems (Poremba et al., 2003; Romanski et al., 1999). There are a number of supporters of the 'what and where' dual-stream processing system (Kaas and Hackett, 1999; Sekiyama et al., 2003). Such a system is required in humans because the auditory system needs to analyse both the identity and location of the incoming stimuli. This is accomplished using the signal from the cochlea and transmitting it through a complex network of pathways and nuclei in the brainstem and thalamus. These structures in the brain then extract spectral and temporal information from the signal in order to calculate the identity and location of the heard stimuli. This dual stream processing model involves a 'what' stream (the ventral aspect which identifies the sound using spectral information) and a 'where' stream (the dorsal aspect which locates and perceives the sound using temporal information). Kaas and Hackett (1999) state that the sensory information is channelled into two parallel pathways which originate from the same cortical areas. These two pathways then diverge into dorsal locations (posterior parietal cortex) and ventral locations (temporal or parietal lobes) (Kaas and Hackett, 1999).

Hickok & Poeppel (2007) also propose a similar **dual-stream model of processing for language**, still using the ventral and dorsal streams. The dual-stream model for language processing also suggests that the ventral stream is bilateral, and helps process speech signals for comprehension. The dorsal stream, on the other hand, is left lateralised and maps speech signals to articulatory networks in the frontal lobes. Hickok & Poeppel (2007) state that the initial acoustic analysis is carried out in the bilateral dorsal superior temporal cortices, including the STG, Heschl's gyri and the planum temporale; this then outputs to the bilateral mid- to posterior- STS where phonetic and phonological processing occur. The STS, in turn, projects to the dorsal and ventral streams - the start of the dual-stream processing system. In terms of the ventral stream, there is processing of distinctive sounds, phonemes, syllabic structures, semantics and syntax. Hickok & Poeppel (2007) propose that this ventral system has its own parallel pathways for the processing of phonological aspects and lexical/semantic aspects separately.

It has been shown that the phonological processing 'sub-stream' is located bilaterally in the primary auditory cortex, but with more left lateralisation shown in both healthy volunteers and brain-injured patients (Belin et al., 2004; Benson et al., 2001; Binder et al., 1997; Binder et al., 2000; Crinion et al., 2003; Hickok, 2009; Hickok and Poeppel, 2007; Miller and Johnson-Laird, 1976; Saur et al., 2008; Scott and Wise, 2003). There is evidence showing left lateralisation of the phonological processing 'sub-stream' in the temporoparietal cortex, the anterior superior temporal cortex and the inferior anterior cingulated cortex. It has been suggested that the left temporoparietal region near to the angular and supramarginal gyri - i.e., the posterior portion of the STG and Heschl's gyrus - play a critical role in phonological coding and therefore in categorical representation (Dehaene-Lambertz et al., 2005); (Liebenthal et al., 2010). Specifically, Zatorre et al. (1996) highlight the role of bilateral anterior auditory regions in the processing of spectral and temporal stimuli. They state that the left anterior auditory region is involved in the temporal processing, while the right homologous area (along with the right STS) is involved in spectral processing. It has further been proposed that the left planum temporal and the posterior STS are key areas in the spectro-temporal analysis of stimuli (Benson et al., 2006; Jancke et al., 2002; Mottonen et al., 2006).

Phonological processing leads to phonological representations, which can then be used in lexical and semantic processing. This dorsal 'sub-stream' is located more widely in the left temporoparietal regions and the ventrolateral prefrontal cortex (Hickok and Poeppel, 2007). Although the specifics of the location of this stream are less clear, there is a consensus that it is based in the posterior middle temporal lobe (Binder et al., 2000; Hickok, 2010; Hickok and Poeppel, 2007). Lesion studies (Bates et al., 2003), fMRI studies (Binder et al., 1997) and PET studies (Fiez et al., 1995) all support the involvement of posterior middle temporal lobe areas in lexical and semantic processing. The middle temporal gyrus and dorsolateral prefrontal lesions correlate with auditory comprehension difficulties (Hickok and Poeppel, 2007); verbal fluency is linked to the arcuate and superior longitudinal fascicules (Bates et al., 2003); and semantic processing is linked to the left STS and middle temporal gyrus (Bates et al., 2003). Semantic processing is also linked to areas outside of the middle temporal lobe including the inferior temporal gyrus, the angular gyrus, anterior and posterior portions of the cingulate gyrus, precuneus and the anterior thalamic regions (Bates et al., 2003). The frontal lobe is also implicated in the lexical and semantic sub-stream of the ventral stream, with Binder et al. (1997) suggesting that it plays the role of "language executive". Binder's fMRI study suggests that portions of the prefrontal cortex (inferior, middle and superior frontal gyri and the anterior cingulated gyrus) are involved. A number of PET studies also suggest the involvement of the inferior frontal gyri and posterior frontal gyri (Demonet et al., 1994; Fiez et al., 1995; Wise et al., 1991).

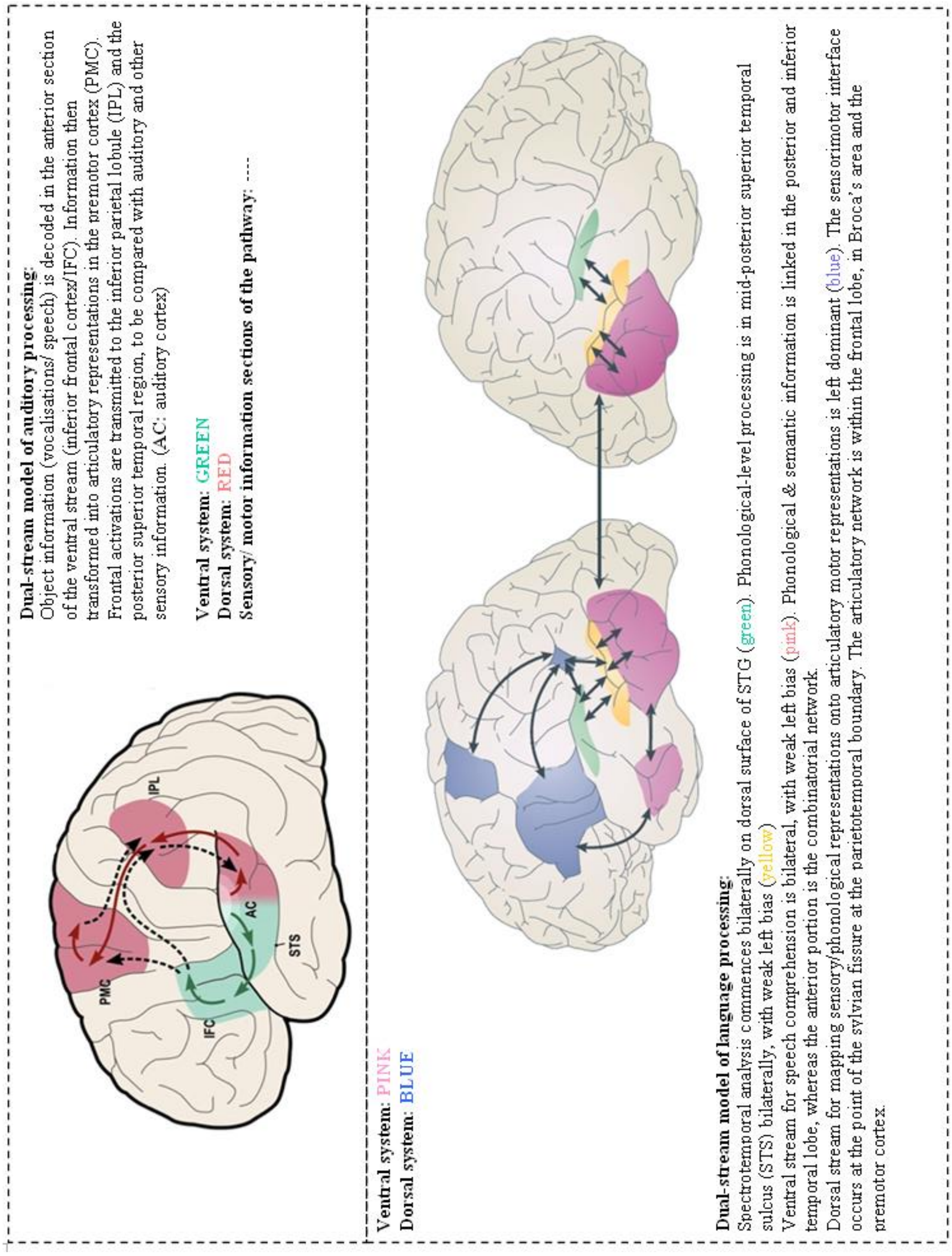

Figure 1-6: Models showing dual streams of auditory processing

Top - Schematic diagram of the dual-stream model of auditory processing (taken from Rauschecker, 2011). *Bottom* - Schematic diagram of the dual-stream model of functional anatomy of language (adapted from Hickok and Poeppel, 2007).

The top row of Figure 1-6 shows a dual-stream model of auditory processing. In comparison, the bottom row shows a dual-stream model of language processing. With the addition of language the model becomes more bilateralised. The language-based model also incorporates the temporal lobe for the processing of phonological and semantic information (posterior and inferior) and for the combination of coded information for speech comprehension (anterior).

A third stream of language processing has been more recently proposed (Scott and Wise, 2004) - the pathway enabling mimicking of sounds. The ‘how’ stream in auditory processing has been proposed as the equivalent to the ‘how’ stream in visual processing, which allows us to reach and grab things in our visual space. It has been suggested to be left lateralised and to run through the supratemporal cortex, including the planum temporale (Scott and Wise, 2004).

Gestural theories (direct realism theory and motor theory) have been investigated in terms of neuroanatomy and a number of fMRI studies demonstrate the role of the motor cortex in speech perception. Sekiyama (2003) found that, when speech is presented, the temporal cortex is activated bilaterally, more specifically the STS and BA22 (overlapping Wernicke’s area). Devlin & Watkins (2007) showed that when TMS is used as a measure of functional connectivity, the motor theory of speech is supported as motor programmes and action words are closely linked. Meister et al. (2007) show the involvement of the pre-motor cortex in speech perception and therefore its role in phonetic perception. These studies are all in agreement with the dual-stream model’s neuroanatomical basis.

1.8. Auditory comprehension

Auditory comprehension, as discussed earlier, is the accumulation of all stages of auditory processing of linguistic components. It is concerned with the nature of incoming auditory signals, as well as the decoding of both the meaning and the underlying concepts that these auditory signals represent. It is the ability to understand auditory messages that allow one to engage in conversations, follow directions and understand messages.

Auditory comprehension involves, and relies on, a number of skills to allow one to fully understand these auditory messages. These skills include auditory memory - the ability to retain auditory information, immediately and following a delay - as well as working memory, i.e., the ability to manipulate this information. Another important skill is auditory closure; the ability to make sense of these auditory messages when sections are missing. Linguistic auditory processing is a term sometimes used to refer to all the skills above, leading to an ability to understand, interpret, organise, manipulate and retain spoken language.

1.9. Cognitive models of auditory comprehension

Auditory comprehension can be broken down into two levels: single word comprehension and sentence comprehension. Comprehension of single words has been modelled far more frequently, and in greater detail, than the comprehension of sentences. Although there are a number of models proposed for sentence production (e.g., (Caramazza, 1997; Dell, 1986; Garrett, 1975), very few have been proposed for sentence comprehension. Most studies, therefore, follow the method described by Saffran (Saffran, 2001), which suggests the use of sentence production models to inform sentence comprehension by working backwards.

In 1980, Morton and Patterson (Morton and Patterson, 1980) proposed the Logogen model of auditory comprehension. This model was discussed in detail by Ellis and Young (2004) and later fed into the proposed fuller cognitive neuropsychological models of language processing[1] The Logogen model (Morton and Patterson, 1980) suggests three main levels of processing for auditory comprehension: first, auditory speech level (auditory phonological analysis); second, lexical level (phonological input lexicon); and third, semantic level (semantic system) (Figure 1-7). It is important to note that, although the authors were aware of the limitations of such a model, it proved to be a worthwhile starting point for the modelling of auditory comprehension.

[1] discussed in detail later in this chapter with regards to impairments associated with aphasia, and used throughout this thesis for purposes of patient classification and levels of breakdown according to impairments.

There have been, and continue to be, great debates regarding a number of features involved in this model. A number of researchers have proposed a less linear model, containing more interaction between levels (Marslen-Wilson, 1987; McClelland and Elman, 1986). The interaction between the levels has also been debated according to the involvement of top-down and bottom-up processes (e.g. Kashino, 2006), as discussed in the earlier section on speech perception. Controversy also exists regarding the early stages of this model, and the topic of specificity or generalisation of prelexical processing is still debated: are there nonspecific acoustic processes carried out on the lexical information or is it speech-specific?

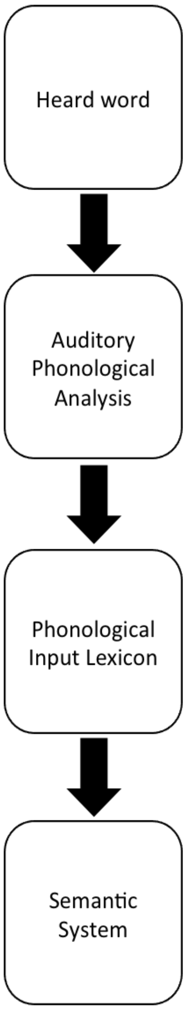

Figure 1-7: Logogen model of auditory comprehension (Morton and Patterson, 1980). (Redrawn from Morris and Franklin, 1995).

Voice perception and speech perception form the basis of auditory comprehension. This thesis therefore uses the dual-stream model of functional anatomy of language

(Hickok, 2009; adapted from Hickok and Poeppel, 2007) when investigating speech perception and the 'auditory face' model (Belin et al., 2004) when investigating voice perception (detailed in Chapter 3). Moving from perception to comprehension requires, among other things, the involvement of the semantic system. Thus, this thesis uses the cognitive neuropsychological model of single word processing to investigate this stage (see Chapter 2 [Figure 2-8] for the adapted and extended cognitive neuropsychological model of single-word processing).

1.10. Neuroanatomy of auditory comprehension

Specht (2014) reviews the neural underpinning of speech comprehension, concluding with the functional neuroimaging support for the dual-stream model of speech perception and comprehension (as well as speech production - not detailed within this thesis). The current model proposes a bilateral ventral stream (containing two parallel 'sub-streams') and a left-lateralised dorsal stream, but does not differentiate between the left and right hemispheric difference in the ventral stream. Specht (2014) proposes the division of the ventral stream into two streams (left and right) to take into consideration voice perception and auditory comprehension. This assumes, therefore, that (i) the left ventral stream is involved in phonetic discrimination, phonological processing, lexical processing, combinatorial processing, as well as semantic processing (i.e., processing the same information as the two parallel 'sub-streams' of the ventral stream within the dual-stream model), and (ii) the right ventral stream is more involved in voice identification and prosody processing. This extends the dual-stream model into a "triple-stream" auditory comprehension model. Specht's proposed model is presented in Figure 1-8 (Specht, 2014).

Auditory comprehension will be discussed throughout this thesis in the light of the models and neural underpinnings of voice perception, speech perception and single word auditory comprehension, as well as by drawing on the information gained from studies on impaired auditory processing disorders. The comprehension of speech has also been discussed in terms of complexity and the fact that it is a multi-faceted process involving other processes alongside auditory processing. It is crucial to take these aspects into account in terms of cognitive modelling and neuroanatomy.

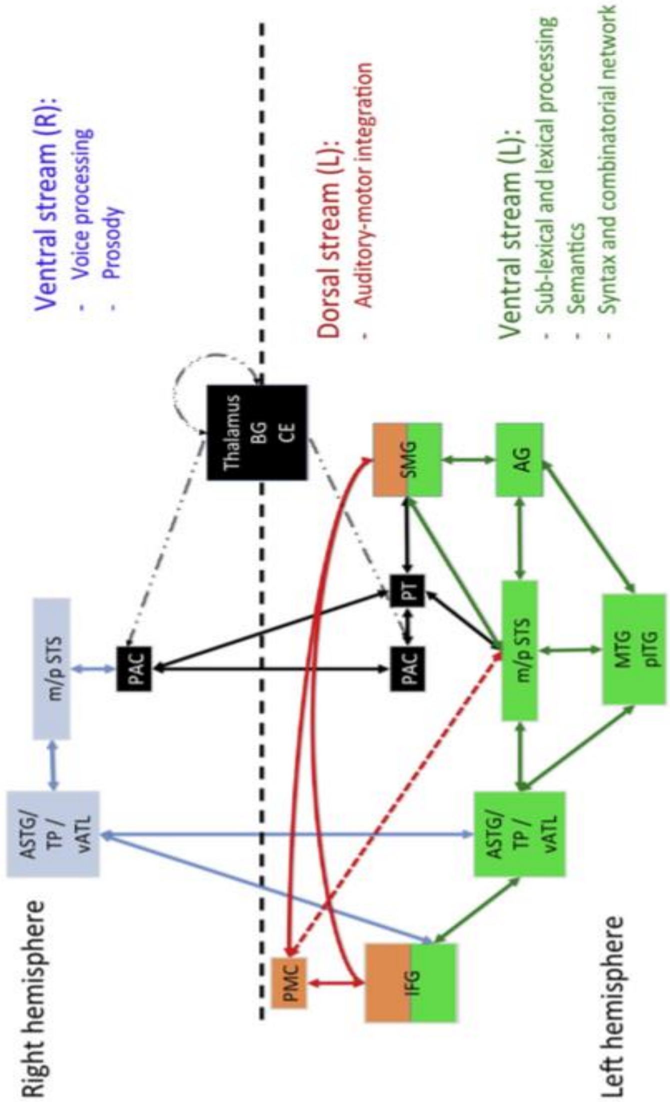

Figure 1-8: Representation of a model of auditory comprehension of speech based on the dual-stream model of processing.

Schematic taken from Specht (2014) showing the triple stream of processing. *KEY:* ASTG (anterior superior temporal gyrus); TP (temporal pole); vATL (ventral anterior temporal lobe); m/p STS (middle/posterior part of superior temporal sulcus); PAC (primary auditory cortex); BG (basal ganglia); CE (cerebellum); PMC (premotor cortex); IFG (inferior frontal gyrus); MTG (middle temporal gyrus); pITG (posterior part of inferior temporal gyrus); AG (angular gyrus); SMG (supramarginal gyrus); PT (planum temporale).

2. Chapter 2 – Introduction to Aphasia

Language is the primary form of communication between humans. An acquired impairment of this language system following a brain injury is known as 'aphasia'. Aphasia can consist of any number of impaired processes in the human language system. This means that aphasia presents differently in every person. It can, therefore, affect an individual's ability to understand, use and/or produce language in part or entirely (Code and Herrmann, 2003).

LaPointe (2005) states that the two main impairments affected in aphasia are spoken output and auditory comprehension. People with aphasia may also have impairments to communication modalities other than speech; gesture, for example, may be affected (Parr et al., 1997). The severity of aphasia also varies from person to person. Some individuals exhibit only occasional word-finding difficulties or auditory comprehension impairment with complex commands in cognitively-demanding situations. Others, however, may be unable to produce spoken output as responsive speech or at the single word level. This severity can also be inconsistent, with many people reporting an effect of fatigue and stress on aphasia severity. Intelligence is not affected in people with aphasia, although a number of cognitive deficits often co-exist (Helm-Estabrooks, 2002; Murray, 2012). Of these cognitive deficits, memory and attention are two of the most prevalent in people with aphasia (Kasselimis et al., 2013; Marshall et al., 2013).

2.1. Classification of aphasia

It is important to note that the studies detailed here all use a specific population of people with aphasia. For the generalisation of results to other populations, it is essential to define the characteristics of the aphasic population being studied (McNeil and Pratt, 2001). The next section looks at the classification systems that have been and/or are currently used in specific aphasia definitions.

Aphasia in people has been documented as far back as 2800 BC (Hatfield and Howard, 1978), although these early cases were pure symptomatic descriptions without any investigation into language theory or neurological mechanisms. The investigation of aphasia in terms of mental processing started in the eighteenth century - with Gesner and Crichton, for example (Eling and Whitaker, 2009) - but it wasn't until the nineteenth century that anatomical and functional debates really dominated. During this period, Paul Broca and Carl Wernicke contributed to the localisation theories of aphasia, identifying and dividing differing aphasia types according to symptoms and location of the lesion in the brain. In the mid-1800s, Jackson (Jackson, 1864) became instrumental in the development of the cognitive school of classification of aphasia, distinguishing between two main speech levels: emotional and intellectual. Following Jackson's cognitive school of thought, Lichtheim (1885) elaborated upon Wernicke's localisation theories and proposed a schematic showing the language centres of the cerebral cortex and their connections, focussing on the functional aspects of aphasia classification (Figure 2-1):

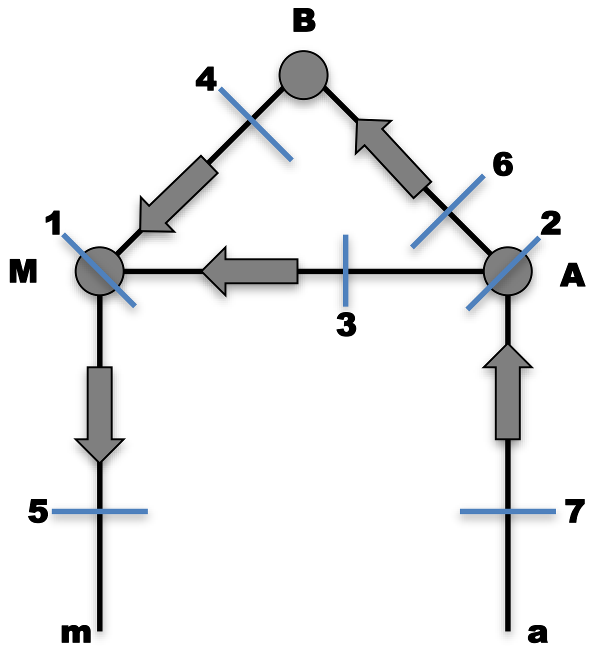

Figure 2-1: Lichtheim's initial schematic of the language system in the brain.
Redrawn from Lichtheim (1885). A: Input store of auditory lexical information; M: Output store of motor word images; B: Concept centre. Black lines indicate connections between centres A, M and B. 'm' is the efferent branch of the connections whilst 'a' is the afferent branch. Arrows indicate direction of flow between centres. Blues lines indicate an 'interruption' of information flow causing an aphasia type. Numbers indicate seven different aphasia types identified by Lichtheim: 1. Motor aphasia; 2. Sensory aphasia; 3. Conduction aphasia; 4. Transcortical motor aphasia; 5. Subcortical motor aphasia; 6. Transcortical sensory aphasia; 7. Subcortical sensory aphasia.

However, Lichtheim was also the first to dismiss his own model and construct a new one. The subsequent model recognised the diffuseness of the 'concept centre' (B), which we now know as the semantic system. In his 1885 paper he also presents an adaptation of his original model showing a number of distributed, but also interconnected 'concept centres' (Lichtheim, 1885) (Figure 2-2).

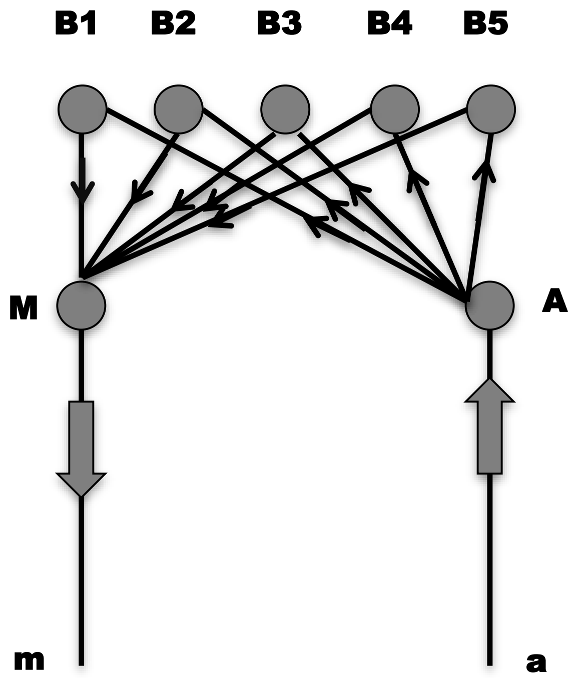

Figure 2-2: Adaptation of Lichtheim's initial schematic of the language system in the brain. Redrawn from Lichtheim (1885). A: input store of auditory lexical information; M: output store of motor word images; B1 – B5: diffuse concept centres. Black lines indicate connections between centres A, M and B's. 'm' is the efferent branch of the connections whilst 'a' is the afferent branch. Arrows indicate direction of flow between centres.

Lichtheim further developed his model, introducing reading and writing, as shown in Figure 2-3.

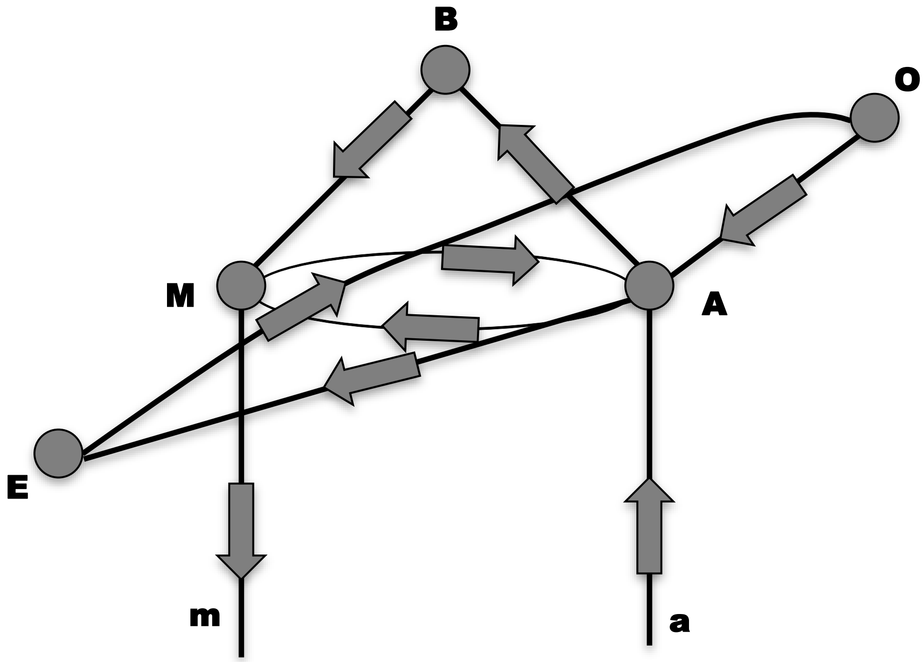

Figure 2-3: The revised Lichtheim model incorporating reading and writing.
Redrawn from Lichtheim (1885). A: input store of auditory lexical information; M: output store of motor word images; B: concept centre; O: reading; E: writing. Black lines indicate connections between centres A, M, O, E and B's. 'm' is the efferent branch of the connections whilst 'a' is the afferent branch. Arrows indicate direction of flow between centres.

Although Lichtheim's models had many shortcomings, they formed the basis for the box-and-arrow cognitive neuropsychological models that are now so familiar in the discipline.

Since the nineteenth century, there have been various other approaches to the classification and division of aphasia types. Marie (1906) proposed that all people with aphasia have a comprehension deficit of some degree and type, but did not elaborate further. Following this, Geshwind (1965) reintroduced Wernicke and Lichtheim's disconnectionist approaches. Geshwind's model uses a schematic representation to show not only that certain areas of the brain are connected, but also that they can be damaged separately. In addition, this model links the visual input system into the previous Wernicke-Lichtheim model (Figure 2-4).

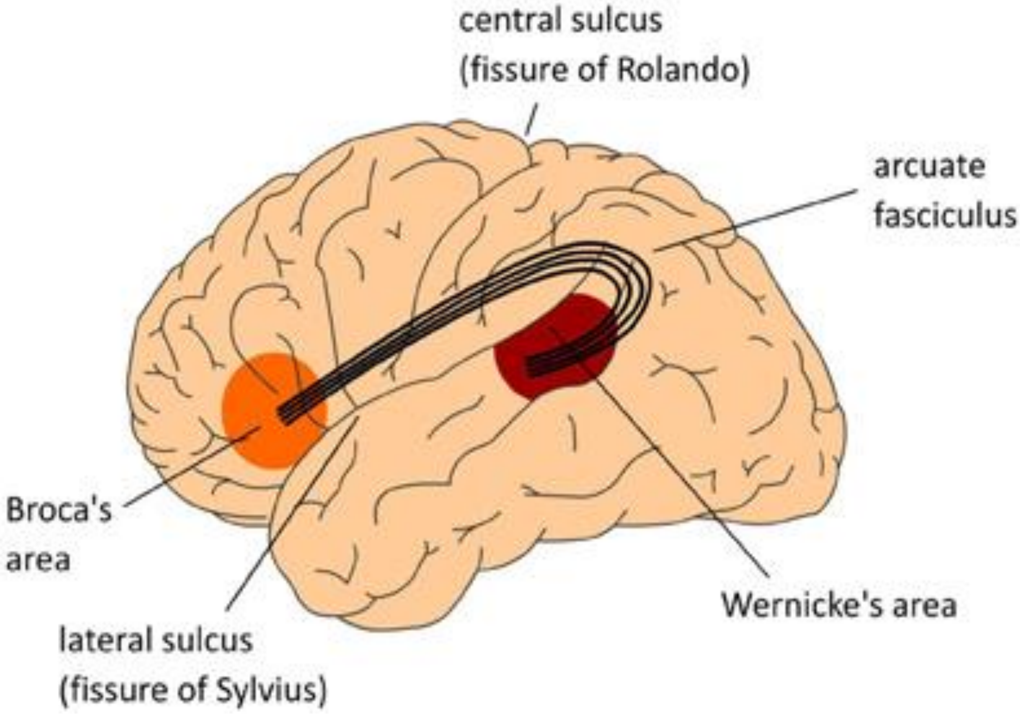

Figure 2-4: Geschwind's model of the neurobiology of language.
This model combines Wernicke's, Lichtheim's and Geschwind's work. Here Broca's area is prominent in language production, Wernicke's area in language comprehension and the arcuate fasciculus connects the two main areas. The arcuate fasiculus is an important bundle of fibres allowing for the exchange of information between these two areas. These language areas neighbour the Sylvian fissure (or lateral fissure) in the brain – an area now known as the perisylvian cortex. (Diagram taken from Hagoort, 2013).

Currently, four main types of aphasia classification systems are in use world-wide. Combinations of two, three or all four are used by many clinicians and researchers:

2.1.1. Boston Classification System:

The Boston Classification of language disorders is one of the most clinically acceptable classification systems, and focuses on the description of aphasia deficits and symptoms. This approach assumes that a specific set of communication impairments can be used to define a syndrome. The modified version of the Boston Classification system (Benson, 1979) divides patients into fluent and non-fluent aphasias, as well as further classification into aphasia syndromes, providing eight different aphasia subtypes in total:

- Broca's
- Wernicke's
- Conduction
- Global
- Transcortical Motor
- Transcortical Sensory
- Isolation
- Anomic

The Western Aphasia Battery (WAB) (Kertesz, 1982) is a comprehensive aphasia assessment battery used regularly by speech and language therapists worldwide. It employs the Boston Classification system, dividing patients according to impairment scores (corresponding to severity), as can be seen in Table 2-1.

	Fluency	*Comprehension*	*Repetition*	*Naming*
Global	0 – 4	0 – 3.9	0 – 4.9	0 – 6
Broca's	0 – 4	4 – 10	0 – 7.9	0 – 8
Isolation	0 – 4	0 – 3.9	5 – 10	0 – 6
Transcortical Motor	0 – 4	4 – 10	8 – 10	0 – 8
Wernicke's	5 – 10	0 – 6.9	0 – 7.9	0 – 9
Transcortical Sensory	5 – 10	0 – 6.9	8 – 10	0 – 9
Conduction	5 – 10	7 – 10	0 – 6.9	0 – 9
Anomic	5 – 10	7 – 10	7 – 10	0 – 9

Table 2-1: Criteria table allowing for practical classification of patients, according to scores on sections of the WAB assessment (Kertesz, 1982), into the main 8 aphasia types as specified by the Boston Classification System (Benson, 1979).
(Redrawn from Kertesz, 1979).

It is important to note that, in practice, this idea of homogeneous groups only works for approximately 30% of aphasic patients (Gordon, 1998); the remaining majority do not fit neatly into these syndrome classifications. This approach is therefore often utilised for educational purposes: to provide clinicians' with an overall understanding of the wide variety and severity of impairments, as well as providing a general picture of the main symptoms that tend to go together. In practice, clinicians often use a combination of classification systems alongside clinical judgement. Nonetheless, this classification system is widely used by researchers who require the recruitment of a specific homogeneous group of patients as well as a standardised, formal classification system. Therefore, this thesis uses the Western Aphasia Battery (Kertesz, 1982) to assess and classify patients according to the Boston Classification System.

2.1.2. Localisationist Model

As stated at the beginning of this section, localisationist models existed prior to the single case studies of Wernicke and Broca. Localisationist approaches, such as Geshwind's adaptation and expansion of the Wernicke-Lichtheim model, use the location and size of the lesion to classify patients. This system is rarely used in isolation; advances in neuroimaging methods have shown variance in symptoms despite similar lesion location and size in the brain (Papathanasiou et al., 2011). This advance in neuroimaging has, however, increased interest in localisationism within current neurology and neuroscience scholarship. As a result, this approach is still in use in research and clinical practice, but almost always in conjunction with another classification system.

2.1.3. Functional Communication Approach

Another major system of aphasia classification, this approach does not classify patients into distinct groups. Instead, this approach focuses on the individual, looking at the effects of the aphasia impairment on all aspects of their real life functioning. It is widely used by clinicians as it can easily be generalised to everyday communication.

It is, however, more difficult to assess and define this approach than the other three approaches discussed here. Consequently, it is used less widely in research.

2.1.4. The Cognitive Neuropsychological Approach

This classification system is employed mainly by clinicians and researchers to study single cases, to further inform knowledge of language systems, and to direct specific targeted impairment-based therapy in clinical practice. This thesis, for instance, uses the Cognitive Neuropsychological Approach, in conjunction with the Boston Classification System, for patient classification in the recruitment and analysis. These approaches are also used as the basis of the single case study in Chapter 5, which looks at a rare case of crossed aphasia.

The cognitive-neuropsychological models of single word processing (comprehension and expression) are static models, which allow an individual's performance to be mapped at a single time-point. There have been a number of models published and used over the years, developing as progress is made within the scientific and clinical worlds. Each model has a number of concepts in common: they all consist of modules (boxes) and processes (arrows), which have a number of assumptions attached to them. Each module is considered to be: domain specific; information encapsulated (i.e. each module operates independently); mandatory (not under voluntary control); universal (consistent across individuals); and innately specified (Shallice, 1984).

The first of these models to be widely published was the 'Transcoding Model' (Figure 2-5) derived by Ellis et al. (1988) and reflects a large body of clinical work carried out previously, including Morton's proposed Logogen model (Morton and Patterson, 1980), detailed earlier. This model shows spoken language processing on the left, written language processing on the right, inputs at the top, and outputs at the bottom.

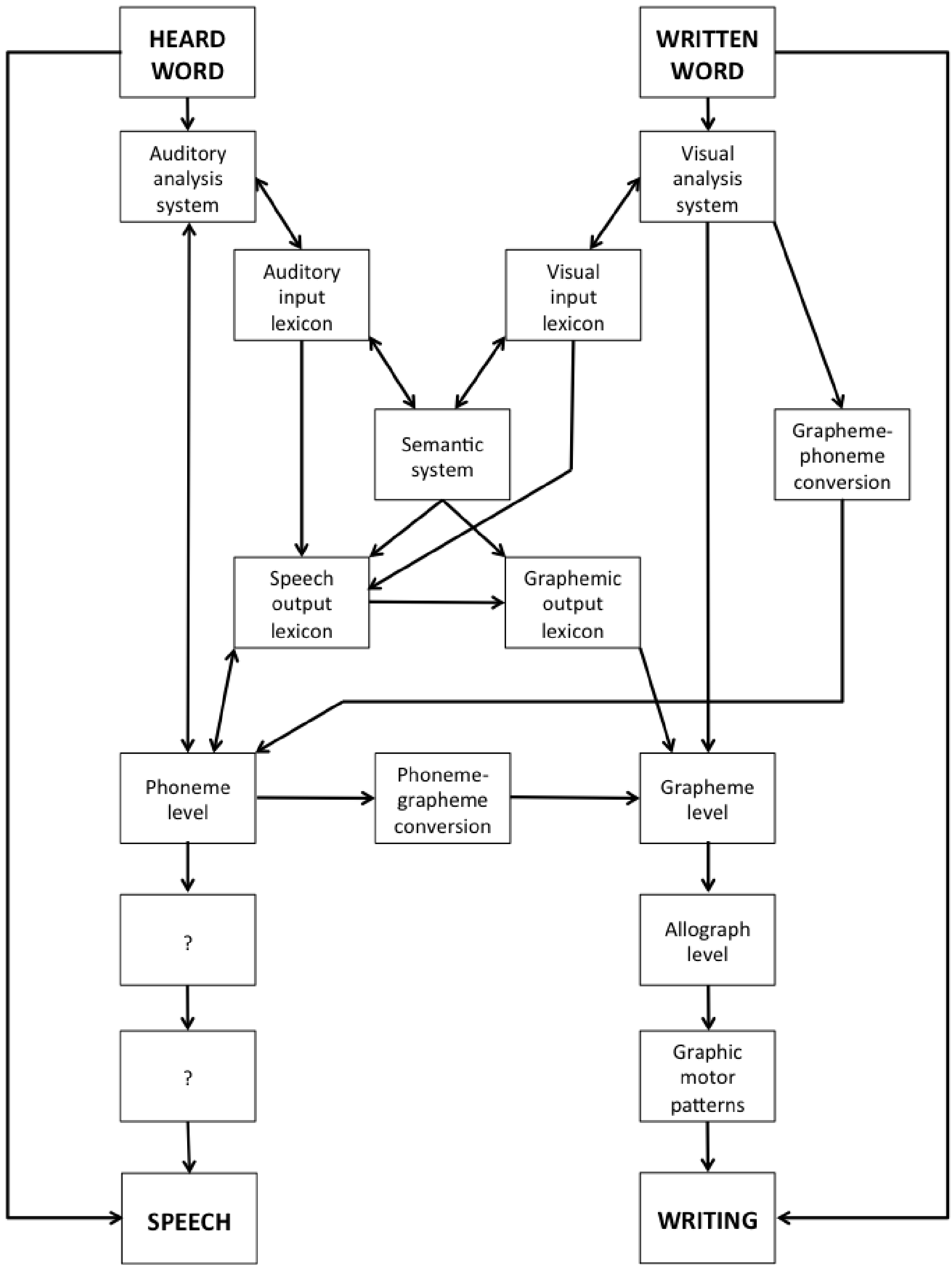

Figure 2-5: Redrawn 'Transcoding Model' originally produced by Ellis et al. (1988).

In 1992, Kay et al. (1992) produced a clinical assessment battery, PALPA (Psycholinguistic Assessments of Language Processing in Aphasia), which was based on, and included, an adaptation of Ellis' (1988) previous model. This model (Figure

2-6) also includes allows for the processing of objects as well as their spoken and written names.

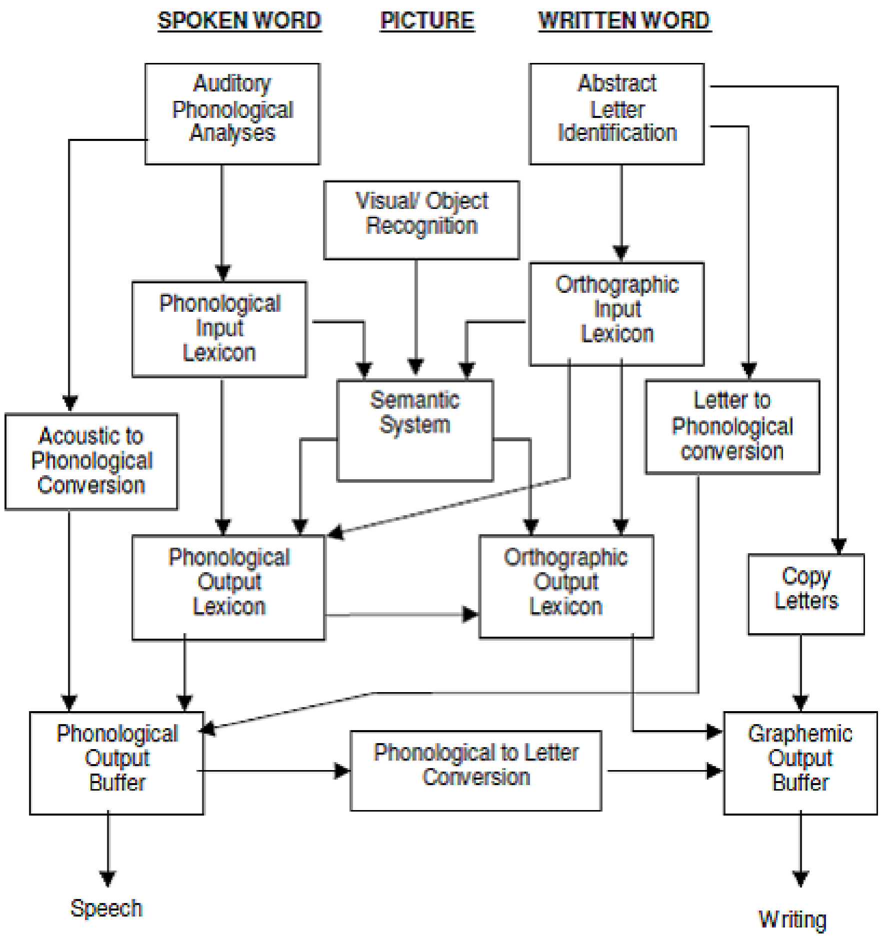

Figure 2-6: Cognitive neuropsychological single word processing model – taken from Kay et al. (1992).

This model (Figure 2-6), used in the PALPA assessment, also includes written and spoken output buffers: Phonological Output Buffer (POB), and Orthographic Output Buffer (OOB). The Phonological Output Buffer is considered to act as a separate temporary storage, or 'hold', for phoneme strings at the point when phonological assembly is taking place (Shallice and Vallar, 1990). Following examples of selective disturbances in reported patients, it has also been suggested that the Orthographic Output Buffer functions as a separate temporary storage for lexical orthographic representations at the point when they are being processed for graphemic assembly (e.g. Goodman-Schulman and Caramazza, 1987; Miceli et al., 1985; Posteraro et al., 1988).

An updated version of the original cognitive neuropsychological single-word processing model has recently been published (Whitworth et al., 2014). This model follows the main principles and structure of the Ellis model (1988), with extensions, but does not include the output buffers included in the model of Kay et al. (1992) (Figure 2-7):

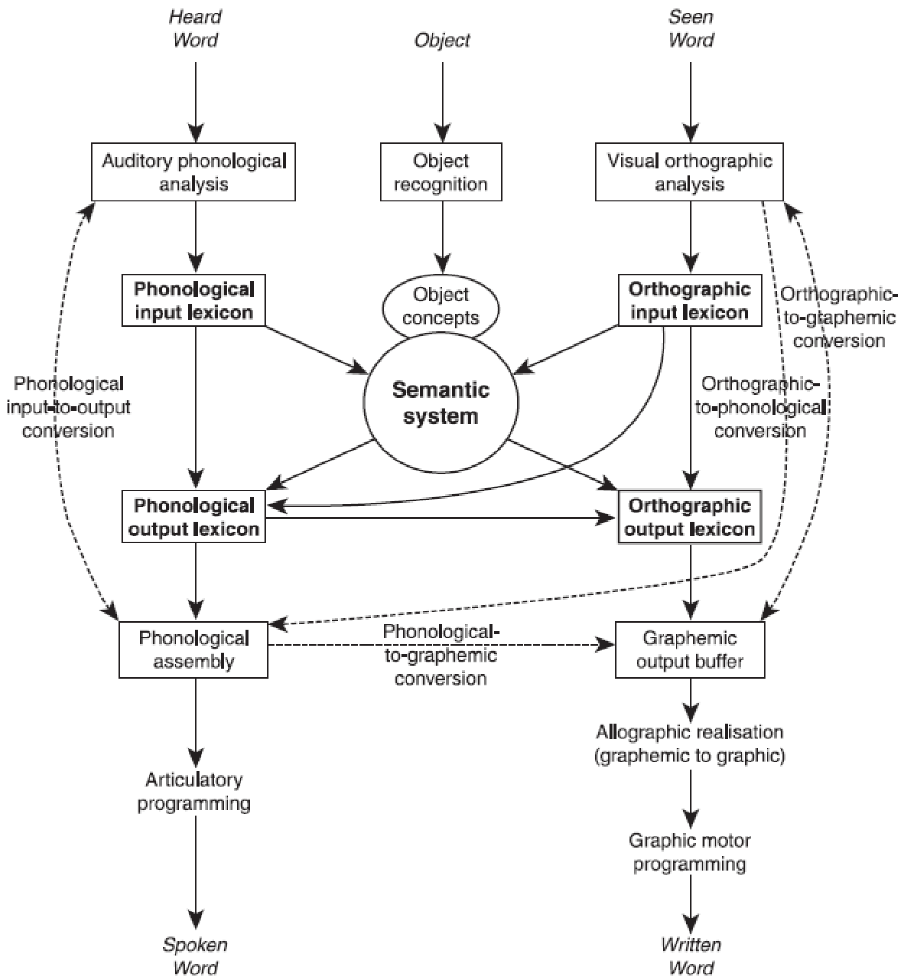

Figure 2-7: Cognitive neuropsychological single word processing model – taken from Whitworth et al. (Whitworth et al., 2014).

This thesis uses a combination of the major published models (summarised above) throughout, and in particular when investigating the single case study (Chapter 5). This combination is an extended version of Whitworth's (2014) adaptation of the Ellis (1988) model of language processing for single words. The extensions are the inclusion of the phonological input buffer (Laganaro and Alario, 2006; Martin et al., 1999; Nickels et al., 1997), phonological output buffer, and orthographic input buffer (Kay et al., 1992), as well as the inclusion of dual processing routes (Ellis et al., 1988) (Figure 2-8).

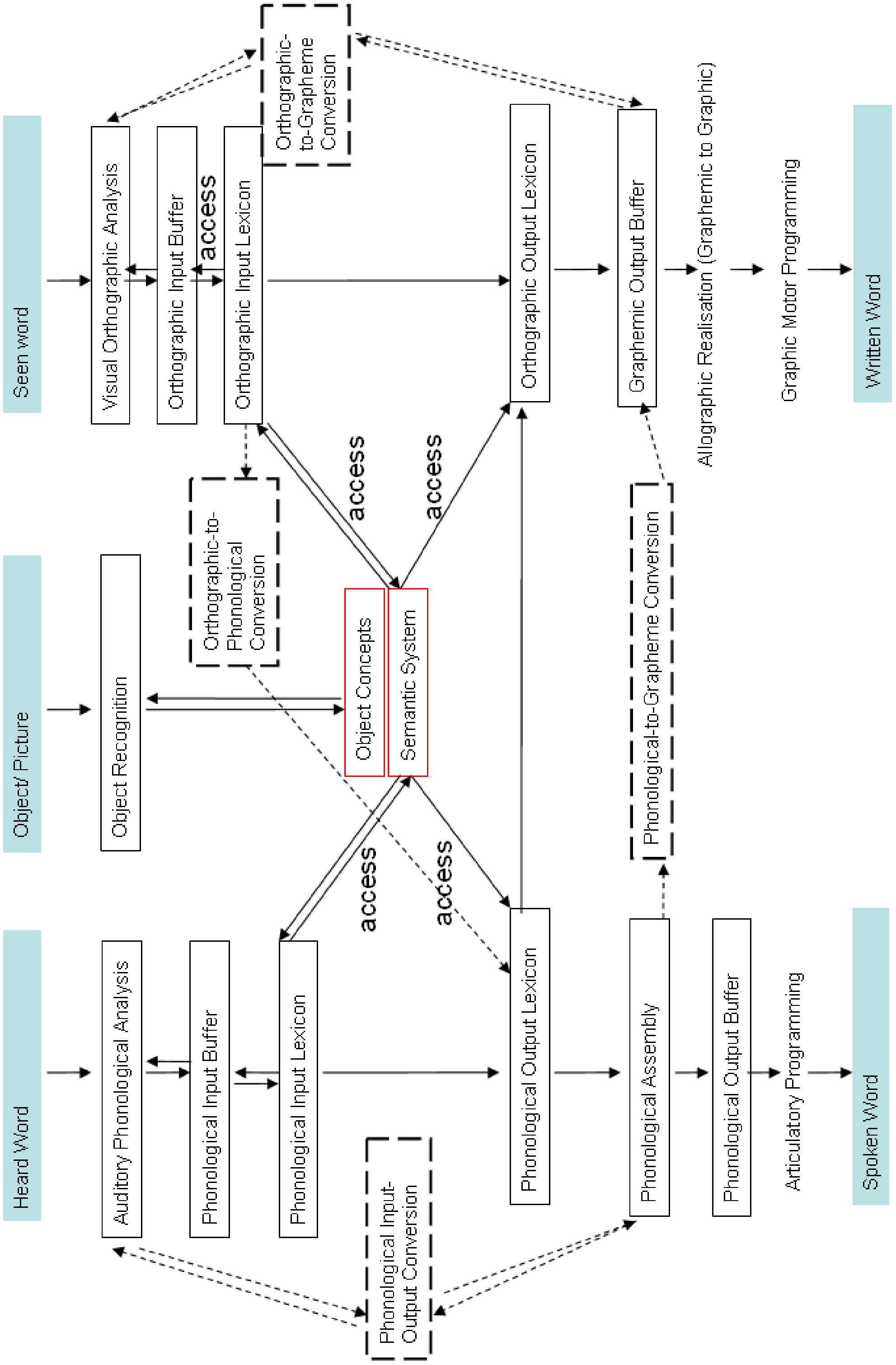

Figure 2-8: Adapted and extended cognitive neuropsychological model of single word processing. Spoken language is represented on the left, written language on the right. Five main processing routes are shown: speech perception (top left); speech production (bottom left); reading (top right); writing (bottom right); and visual object input (top centre). The semantic system (along with the object concept

system) is the main commonality to all routes (centre, red). Boxes represent independent modules; arrows represent flow of information (in the direction of the arrow) between the specified processor/module; blue rectangles represent input and output modalities; dashed lines represent information flow that is bypassing modules; dashed boxes represent modules which are involved in this 'bypass' process; "access" represents flows of information that can become impaired.

2.2. Brain damage related to aphasia

A cerebral vascular accident (infarct or haemorrhage) is the most common cause of aphasia (The Stroke Association, 2008). In the majority of cases, the lesion is located within the left hemisphere. Indeed, in the normal population, most individuals show 'language centres' located in the left side of the brain: 90% of the population is right handed and 75% of right handed subjects show strong left language lateralization, while more than 90% of left handed subjects show left and non-lateralized language processing (Mazoyer et al., 2014). Manual dominance should therefore not be taken as a strong indicator of lateralization; it only weakly correlates with language lateralization, and is at best partially pleiotropic (Ocklenburg et al., 2014). Early investigations into aphasia anatomy relied on post-mortem analyses (Alajouanine et al., 1939; Broca, 1863; Wernicke, 1874). But even at that relatively early stage, it became apparent that the relationship between lesion site and aphasia presentation is not simple or 'one-to-one'. Advances in imaging techniques have meant that studies of both healthy participants and those with lesions can provide invaluable insights into the localisation of language in the brain (Awad et al., 2007; Dronkers et al., 2007; Fridriksson et al., 2007; Wise et al., 2001). It is now known that there are many brain areas that are involved in language comprehension and production beyond just the classical Broca's and Wernicke's regions.

A large number of studies have investigated the link between lesion location and aphasia type (Awad et al., 2007; Basso et al., 1985; Basso et al., 1995; Bonilha and Fridriksson, 2009; Damasio and Damasio, 1980; Dronkers et al., 2007; Fridriksson et al., 2007; Hillis et al., 2004; Kreisler et al., 2000; Newhart et al., 2007; Willmes and Poeck, 1993; Wise et al., 2001). There remain, however, a large number of discrepancies in the literature as to the consistency and correlation of symptom-lesion mapping. Kreisler et al. (2000) suggest a strong link, whereas Basso et al. (1985) suggest no reliable correlation between lesion location and aphasia type. Kreisler et al. (2000) carried out a classification study looking at the anatomy of aphasia and concluded that, at the acute phase, the main determinant of aphasia type is the lesion location. This study concludes that between 67 and 94% of participants taking part could be classified correctly according to the anatomical location of their lesion. Fridriksson (2013) offers a more recent study linking anatomical lesion location to aphasia type. The paper states that damage to the anterior arcuate fasciculus is a predictor of non-fluent aphasia. However, several studies have also shown that a number of different damaged brain regions can be the cause of a given language impairment (Bonilha and Fridriksson, 2009; Hillis et al., 2004; Newhart et al., 2007). This is due to the fact that the lesion does not impact solely on the proximate region but also on connected regions within the network. There are a number of studies that show extreme deviations such as global aphasia in small, restricted locations; lesions within language areas that do not show symptoms of aphasia; and posterior lesions with accompanied non-fluent aphasia type (Basso et al., 1985; Basso et al., 1995). Therefore it is not possible to accurately infer lesion location from aphasia type with any consistency, nor vice versa. Basso (1995) goes so far as to suggest this methodology is unreliable in at least a quarter of cases.

Table 2-2, compiling information from the WAB (Kertesz, 1982), Kreisler et al. (2000) and Love & Webb (1996), shows the most common lesion sites (according to the included studies) that are associated with each aphasia type detailed in the Boston Classification System (Benson, 1979). As discussed earlier, localisationism is now rarely used as a solitary classification system, due to advances in neuroimaging methods which show that lesions with similar location and size in the brain can and do produce highly varied symptoms (Papathanasiou et al., 2011). When looking at the correlations between lesion location and aphasia type, it is also important to take into consideration the time post-stroke; these anatomical correlations become blurred and less specific with diaschisis and as recovery advances (Godefroy, 2013).

It is still useful, however, to look at the most common lesion sites associated with specific aphasia types when carrying out any imaging and lesion-based study on people with aphasia. This is especially true when using the Boston classification system to group patients behaviourally. Importantly though, these anatomical correlations are not used as a gold standard throughout the studies presented in this thesis, but used as a guidance tool.

At this stage, it is important to recognise that, in practice, it is uncommon for patients to fit neatly into the above anatomical and behavioural classification types. However, in group studies the aim is often to work with a homogeneous group of patients to allow improved understanding and description of that particular population (Goodglass et al., 2001). This thesis uses the aforementioned classification systems, especially the Boston Classification system and the localisationist model, also presented in this chapter. The Boston system is used to group patients according to behavioural presentation. In terms of anatomical classification, once again the formal localisationist system is not used for inclusion/exclusion criteria. Instead a broader system, based on arterial territory is used throughout this thesis. Therefore this thesis aims to investigate patients with lesions who have an auditory impairment following a stroke, and to classify the patients who are recruited into the study to inform the interpretation of the results.

Aphasia type		Disturbances	Common lesion site
Aphasias of the perisylvian zone (cortical aphasias)	Global	Disruption to all language functions	Extensive anterior-posterior lesions of perisylvian association cortex of the left hemisphere
	Broca's	- Speech output is impaired (slow and effortful, missing content words, poor grammar, mis-articulations). - Comprehension only slightly impaired, if at all.	Left frontal lobe, in the posterior third of the inferior frontal gyrus (IFG) frequently extending to the subcortical or insular regions
	Wernicke's	- Speech sounds fluent but lacks meaning, comprises jargon and inappropriate words, inability to name. - Comprehension (particularly auditory) impaired.	Junction between the temporal and parietal lobes, specifically in the posterior STG and in the white matter underlying Wernicke's area
	Conduction	- Spontaneous speech impaired, often contains phonemic paraphasias. Repetition impaired. - Comprehension only slightly impaired, if at all.	Arcuate fascilus and/or the cortico-cortical connections between Wernicke's & Broca's areas
Transcortical aphasia's outside the perisylvian zone (subcortical)	Isolation / mixed transcortical	- Spontaneous speech impaired. - Repetition preserved. - Comprehension impaired.	Perisylvian association cortex
	Transcortical Motor	- Spontaneous speech impaired (effortful and slow). - Comprehension relatively preserved. -Reduced initiation	White matter tracts deep to Broca's area connecting to parietal lobe or frontal lobe (anterior/superior to Broca's area)

	Transcortical Sensory	- Speech relatively unimpaired. Repetition intact. - Single word comprehension impaired, with word meanings disturbed.	White matter tracts connecting parietal lobe to temporal lobe - lesion isolating perisylvian speech structures from posterior brain
	Anomic	- Production in single words and nouns impaired. Repetition intact. - Comprehension intact.	Angular gyrus in parietal lobe or the connections between the parietal and temporal lobes

Table 2-2: The most common impairments and lesion sites associated with each of the aphasia types at the acute stage of recovery.
Table compiled using the classified in the Boston Classification System (Benson, 1979), and information taken from (Kertesz, 1982), Kreisler et al. (2000) and Love & Webb (1996).

2.3. Incidence

As stated earlier, stroke is the main cause of aphasia. Approximately 152,000 people in the UK have a stroke every year (Townsend et al., 2012). Figures for the incidence of aphasia in the acute stroke population vary hugely, with figures ranging from 15% (Intercollegiate Stroke Working Party, 2012) to 25% (Wade et al., 1986). The Intercollegiate Stroke Working Party (2012) estimates that one third of these stroke patients are then left with a chronic communication disability. It is also approximated that at any one time in the UK there are around 250,000 people who have aphasia, many of whom are under retirement age (Law et al., 2007). Therefore, although all figures are approximate and vary, it seems reasonable to conclude that post-stroke aphasia is not a rare impairment.

2.4. Impaired auditory processing

Auditory processing impairment can range from a general hearing impairment to specific problems in auditory comprehension. As yet, there is no clear taxonomy of disorders of auditory processing. This is possibly due a lack of consistency across the terminology used. Disorders of auditory processing can range from cortical deafness to auditory agnosia, from pure word deafness (speech agnosia) to phonological retrieval disorders, and from central auditory processing disorders (CAPD) to aphasia. CAPD in an umbrella term for a number of disorders that involve auditory processing impairments, leading to difficulties with recognising and interpreting sounds, and in particular speech sounds. This term can be further split into two areas: acquired and developmental. Both types of CAPD can be caused by damage to or dysfunction of the central auditory nervous system, often with an unknown underlying cause (Lew et al., 2007; Musiek et al., 2011). However, the exact definition and diagnosis of CAPD still varies (Miller and Wagstaff, 2011). The British Audiology Society (Alles et al., 2011) provides a broad definition, stating that CAPD may result from ear infections, head injuries or neurodevelopmental delays that affect processing of auditory information. Therefore, if using this definition, CAPD can be seen as a component of some subtypes of aphasia.

2.5. Impact of aphasia and an associated auditory processing impairment

Aphasia has a "devastating impact on the lives of people who are – prior to onset – typically fully competent communicators"(Steele et al., 2003 p.98)

Aphasia is often referred to as the 'hidden disability' due to the fact that it is not an obvious outward physical disability, and therefore it does not fit neatly into society's perception of a disability (Parr et al., 1997). However, this does not minimise the impact of this disability on the individual:

"In my head I can talk, only the mouth will not"
(Christensen, 1997 p.733)

The most obvious impact on an individual is the reduced ability to communicate in an effective manner. However, this effect on abilities often leads to increased frustration and social isolation as well as a breakdown in relationships (Speakability, 2006).

Sarno (1997) discusses four major domains involved in the assessment of quality of life for an individual: emotional functioning; social functioning; activities of daily living; ability to actively engage in pleasurable pursuits. When an individual's communication abilities are no longer intact, it becomes a challenge to develop and maintain social relationships (Code and Herrmann, 2003) as well as to participate in many every-day activities. Therefore aphasia can, and does, impact upon all aspects of an individual's life, affecting all four of the domains that Sarno (1997) considers to be essential in assessing quality of life..

Many studies have reported the impact of aphasia on the quality of life. Hinckley (1998) suggests that, in the stroke population, chronic aphasia impacted on quality of life perception more strongly than a persisting motor impairment. Parr (2007) states that frustration and tension are both increased in those with aphasia due to the breakdown of close family relationships. Parr (2007) also looked at the effect of aphasia on participation in society and found that it was rare for patients to return to work at the same professional level as they did prior to the event. Hinckley (2002) also discusses the negative effects of aphasia on participation in social activities. There are many studies that discuss one of the most pertinent effects of aphasia: psychological and emotional wellbeing. Brain injury itself can predispose an individual to anxiety, neglect, depression and emotional lability (Tanner, 2003). These emotional difficulties can further exacerbate the 'grieving process' which individuals with aphasia go through when coping with not only a loss of speech, language and communication skills, but a loss of associated factors such as independence, relationships, control, motivation and self-worth (van der Gaag et al., 2005). A factor compounding all of the above is the increased vulnerability faced by individuals with aphasia due to an inability to seek help or consent to therapy/treatment (depending on the severity of the impairment). Bakheit sums this up, stating that "…aphasia has a significant negative impact on the patient's well-being, independence, social participation and quality of life and is often associated with severe depression" (2007, p.885).

The severity of the communication impairment is not necessarily the main factor contributing to a reduced quality of life for all people with aphasia (Herrmann and Wallesch, 1990); (Sarno, 1993). Ross & Wertz (2003) identify three areas affecting quality of life in a person with aphasia: 1) level of independence, 2) social relationships, 3) environmental factors such as transportation and information accessibility. However, it is clear that many of these factors are actually impacted upon as a direct result of an individual’s impaired communication skills and the severity of the aphasia (Ellis-Hill and Horn, 2000).

When discussing the impact of aphasia, it is also important to note that such communication impairments also impact severely upon family members, friends and carers. Interpersonal relationships are affected, time pressures, and responsibilities are increased, and ways of life are altered (Enderby et al., 2009; Lincoln N et al., 1984).

Looking further afield it is pertinent to state that aphasia also has a negative impact upon society in terms of economic pressures (i.e. such as cost to the patient, carers, families, health service and society), although data pertaining to this is not readily available at present (Brady et al., 2012). Parr (2007) discusses the practical impact of aphasia in that families with two income-generating adults prior are forced to adapt to living on government assistance as the affected individual is unable to work (due to his/her disability) and the partner is also unable to work (due to increased caring responsibilities). The impact upon health services around the world is both long term and wide reaching. Speech and language therapy can be required for an individual with aphasia, depending on severity and a number of other factors. This therapy is often required across all stages and phases of recovery for both the individual and others involved in their lives – often for education and training purposes. This is especially important when other professionals are involved in the care of the individual with aphasia - if they are unable to communicate effectively with their client then intervention and support could be ineffective and inappropriate (Law et al., 2007).

Aphasia has been shown to have a huge and widespread impact not only upon aspects of an individual’s emotional, social, environmental, and day-to-day life, but also on the lives of family members, friends and carers, and on society in general, particularly in terms of economic pressures.

Therefore, the need to undertake further research into this debilitating and devastating condition is clear. In addition to the aim of investigating auditory comprehension, this thesis contains a pilot study that examines a novel computer-assisted therapy that can be performed at home, thus possibly providing cheap and readily available therapy for many people.

3. Chapter 3: Intact Voice and Speech Perception

3.1. Perception of voices

There are many factors involved in the perception of vocal sounds, both linguistic and paralinguistic, including amplitude, pitch, and timbre. Amplitude (a measure of how much energy a sound wave has) is perceived as volume and can be visualised as the height of a pure tones on a sine-wave (Figure 3-1). Pitch is related, mainly, to the perceived fundamental frequency (fo) of the vocal fold vibration, i.e. the frequency with which the waveform repeats itself during speech. However, pitch is not a single dimension running from perceived 'high' pitch to perceived 'low' pitch. It consists of two dimensions: pitch chroma and pitch height, often represented as a helix with pitch chroma as the circular dimension and pitch height as the vertical dimension, as seen in Figure 3-2 (Warren et al., 2003). A simple way to understand the difference between pitch chroma and pitch height is through notes and octaves on a keyboard. Pitch height can be illustrated by an octave, and pitch chroma by the cycle of notes between the octaves. Timbre (seen in the power spectrum of the sound) is perceived as the characteristic quality of a sound, and is determined by the relative strengths of the different frequency components, which in turn are determined by the resonance (increase in amplitude when a sound wave collides with a structure in its path) (Hewlett and Beck, 2006). Timbre is directly influenced by physical factors such as age and gender (Belin et al., 2004) and is often described as a 'holistic' attribute of sounds (Wile and Balaban, 2007).

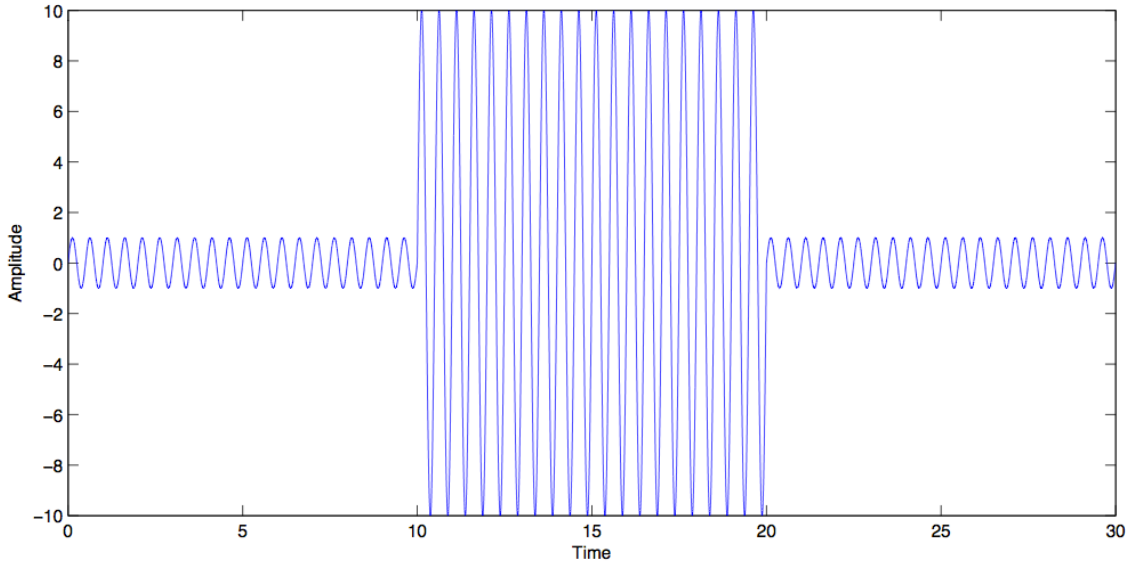

Figure 3-1: Sinewave with varying amplitude.

Centre shows greater amplitude, with reduced amplitude either side.

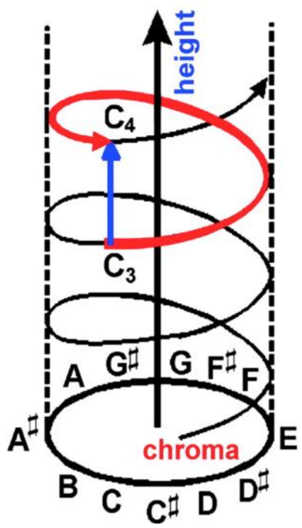

Figure 3-2: Pitch as two dimensions through the illustration of a musical scale.

Musical scale is circular with each circuit representing an octave (red). Pitch height is shown as the equivalent change of a circular octave via a vertical line (blue). Taken from Warren et al. (2003).

Perception of paralinguistic information in a voice has only recently been studied. Voice perception implies drawing out information from the actual voice of an individual, something used on a daily basis to determine the emotions, age and gender of the speaker. Therefore the voice does not only contain speech information, but holds a whole host of other information too. Many researchers in the field now refer to it as the "auditory face" (Belin et al., 2002; Belin et al., 2004; Belin and Grosbras, 2010). This term refers to the fact that voices allow us to recognise an individual and their emotional state. There is an overall structure that carries detailed information that varies, combining to form a voice with three main sections: speech information, vocal affective information, and vocal identity information (Ellis et al., 1988). Voice contains information that provides an ‘identity’ in the form of features such as the timbre or the accent. It also provides information regarding the affective state of the speaker i.e. their emotional status, via amplitude, f0 variation, duration of pauses, and other features (Belin et al., 2004). Of particular interest for this thesis, specific vocal features can contain more than a single type of information, for example formants carry speech information (phonetic information) (Klatt, 1980) as well as identity information (vocal tract size and the directly related body size) (Fitch, 1997).

Gender is an important part of the identity information present in our voices. Infants can recognize male/female voices from an early age, signifying the importance of voice gender perception (Mullennix et al., 1995). It is the first physical factor we determine through voices, and healthy listeners have been shown to extract gender information from the voice relatively easily and accurately (Bachorowski and Owren, 1999; Garrido et al., 2009; Mullennix et al., 1995). Gender perception in voice is affected by the size of the larynx and vocal tract (Belin et al., 2004; Lass et al., 1976) and is perceived using both pitch (Landis et al., 1982) and timbre (Bachorowski and Owren, 1999). Pernet & Belin (2012) investigated the relationship of pitch and timbre in gender voice perception and state that pitch and timbre are used commonly, in conjunction, to determine voice gender. However, pitch ranges, determined by formant fundamental frequencies, can overlap in males and females making the use of pitch alone in voice gender categorisation unreliable at times (Hanson and Chuang, 1999). It has been suggested that timbre is ‘integral to recognition mechanisms’(McLachlan and Wilson, 2010) and Pernet & Belin (2012) showed that it is actually possible to categorise voice gender from timbre alone. The study presented in this chapter and in chapter 4, use gender as the paralinguistic feature for investigating voice perception. It is important to note that gender was chosen as it is signalled through both pitch and timbre, although another paralinguistic factor could have been used.

3.2. Cognitive models of voice perception

The field of voice perception remains much less advanced than the field of speech perception. However, models have been formulated recently that are strongly influenced by the models of face processing, with parallels between the two domains proposed. The face recognition model proposed by Burton et al. (1990) has been the basis for the Ellis et al. (1997) model of voice recognition. This model suggests that firstly 'auditory structural encoding' occurs, when vocal feature information is encoded at a very basic level. This information is then further processed and put together in 'voice recognition units'. There are 'person-identity nodes' also present which link with the 'voice recognition units' to enable us to make decisions regarding familiarity of voice. Lastly 'semantic information units' provide the semantic information about the incoming voice.

More recently Belin et al. (2004) proposed a model of voice recognition that built upon this initial model, also including information regarding the processing of other vocal information, and so the idea of the 'auditory face', discussed earlier. It is also based on Bruce and Young's model of face perception (Bruce and Young, 1986) and includes both cognitive and perceptual aspects of processing. The model (Figure 3-3) shows the processing of voices firstly at a general low-level along with other auditory stimuli. This is followed by a 'structural encoding' stage and then parallel, independent, processing systems that analyse the three aspects of voice processing mentioned above: 1) speech information; 2) vocal affective information; 3) vocal identity information. The model proposes that these three processing pathways interact with similar pathways within the face-processing model. However, Belin et al. (2004) suggest that these three pathways are not completely independent until they reach the highest levels. Belin et al. (Belin and Grosbras, 2010) also stress, in a later paper, that the proposed model does not conclude that all stages of voice processing are analogous with face processing, but instead the model can be used as a guide for further research in this field and as a basis on which to build future models.

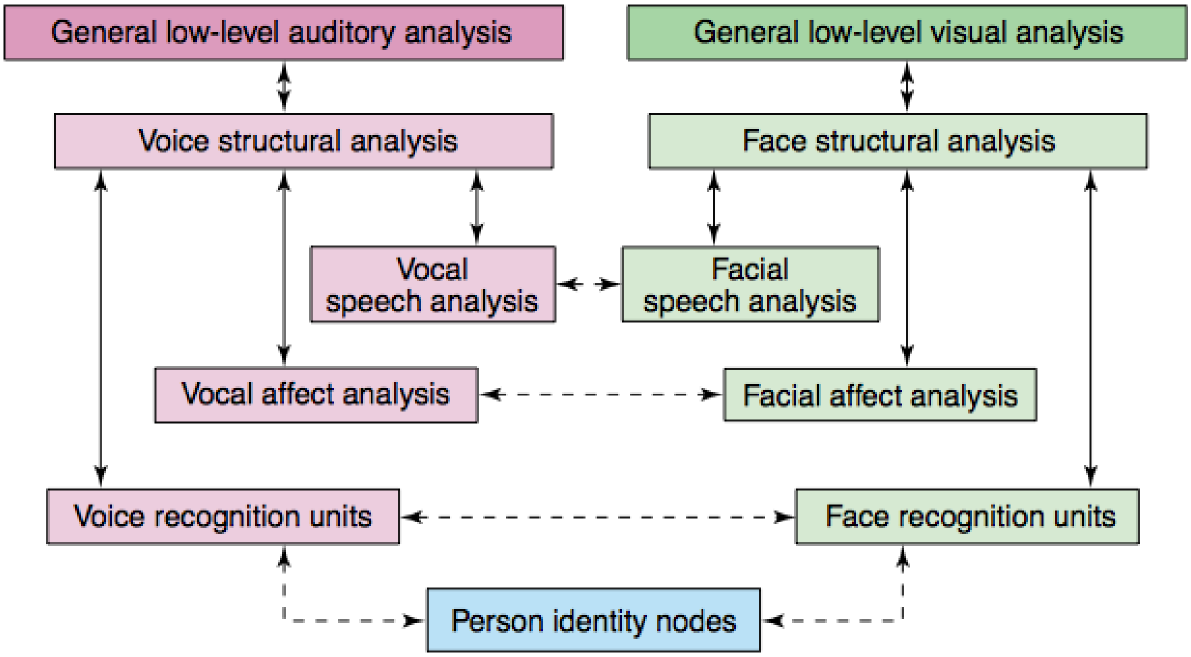

Figure 3-3: A model of voice perception.

Taken from Belin et al. (2004). Right: originally adapted from Bruce and Young's model of face perception (Bruce and Young, 1986). Left: voice processing (Belin et al., 2004).

3.3. Voice perception neuroanatomy

The neuro-anatomical basis of voice perception has been less extensively investigated than for speech perception. The debate over whether voices are 'special' or not remains unresolved i.e. does voice perception involve specialised modular brain mechanisms or not? Similar debates are ongoing in many other areas of perception research e.g. speech perception (Kluender et al., 2006; Liberman and Mattingly, 1989; Liberman and Whalen, 2000) and face perception (Gauthier and Bukach, 2007; McKone et al., 2007).

Bilateral voice-sensitive regions (Pernet et al., 2015b) with an increased sensitivity in the right hemisphere have been found in a number of fMRI and transcranial magnetic stimulation (rTMS) studies comparing vocal to non-vocal responses (Belin et al., 2002; Benson et al., 2001; Celsis et al., 1999; Lattner et al., 2005; Mummery et al., 1999). Temporal Voice Areas (TVA) were first proposed by Pernet et al. (2007) with greater activation shown bilaterally in the superior temporal sulcus/gyrus (STS/STG) in response to vocal than non-vocal sounds (Ahrens et al., 2014; Belin and Grosbras, 2010; Pernet et al., 2015b). Three specific voice-sensitive areas have been localised along the posterior, mid and anterior STS/STG (TVAp; TVAm and TVAa) bilaterally (Figure 3-4) (Ahrens et al., 2014; Pernet et al., 2015b). It is still not clear the exact function of these areas but it has been postulated that both the TVAa and TVAm are linked to voice-specific acoustical processing (Charest et al., 2012) whereas the TVAp is more involved in the integration of audio-visual stimuli (Watson et al., 2014). Extra-temporal areas have also been identified as voice-sensitive, including the bilateral inferior prefrontal cortex and amygdalae (Pernet et al., 2015b). Again, the specific role of these areas is not clear, but it is known that the amygdalae is involved in affective processing as well as the relevance of vocal output in the human environment (Pernet et al., 2015b).

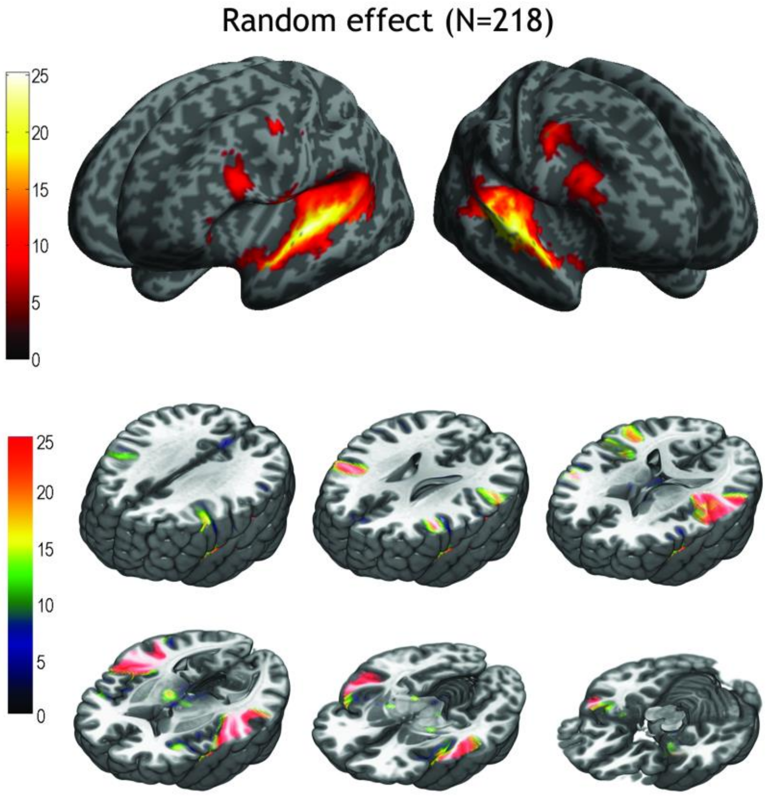

Figure 3-4: Activations of the temporal and extra-temporal voice-sensitive areas during passive listening to vocal and non-vocal stimuli

Random effects analysis (RFX) from n=218 subjects showing significant values ($p<0.05$, FWE corrected) for vocal vs. non-vocal stimuli during a voice-localiser fMRI study. Taken from Pernet et al. (2015).

A number of neuroimaging studies have been carried out over the last decade investigating specific aspects of voice perception. It has been suggested that the bilateral anterior STS regions that are sensitive to vocal sounds (Belin et al., 2000) have different preferences laterally: the left STS regions are responsive to linguistically-based information and the right STS has a preference for paralinguistic information (Belin et al., 2002; Celsis et al., 1999; Crinion et al., 2003; Scott and Wise, 2004).

The neuroanatomy of the perception of voice-identity is still relatively unknown but the bilateral anterior temporal lobes (towards the temporal pole) have been linked to the perception of voice identity, regardless of familiarity. In terms of familiarity of voices, the superior temporal pole and the inferior frontal cortex (both bilaterally) are more highly activated with unfamiliar voices (Latinus et al., 2009), whereas the inferior temporal pole, the frontal parietal cortices, the posterior superior temporal cortex and the fusiform gyrus (all bilateral) are more highly activated with familiar voices. Lattner et al.'s (2005) study looked specifically at gender and pitch perception in the right hemisphere, and state that voice spectral information is processed bilaterally in posterior parts of the STG and in the areas surrounding the planum parietale (i.e. the superior extension of the sylvian fissure).

Gender perception is an important part of voice identity perception. Charest et al. (Charest, 2009) confirms that voice gender categorisation occurs in the anterior part of the right STS, using a higher level network (including bilateral frontal gyrus, prefrontal cortex, insular and anterior cingulated cortex) to bring together all the acoustical features from the auditory areas (Belin and Zatorre, 2003). Charest et al. (2012) also confirms the role of the right hemisphere and in particular the right frontal cortex in voice-gender processing (Mullennix et al., 1995).

'Affect perception' is another aspect of overall voice-perception, and one that has been investigated for a long time. Early clinical work (Heilman et al., 1984; Hornak et al., 1996; Ross, 1981; Van Lancker and Sims, 1992) suggested that lesions in the right hemisphere were associated with impaired recognition of vocal affect than lesions in the left hemisphere. These results were confirmed by more recent neuroimaging studies (Buchanan et al., 2000; Ethofer et al., 2007; George et al., 1996; Mitchell et al., 2003; Morris et al., 1999; Rama et al., 2001; Wildgruber et al., 2000) which have shown that the bilateral temporal cortices are involved in vocal affect perception, with more lateralisation in the right hemisphere. Specifically, Ethofer et al. (2011) located Emotional Voice Areas (EVA) as situated postero-laterally to the primary auditory cortex in Heschl’s gyrus and extending along the STS/STG, in both hemispheres but more right lateralised.

To summarise, two main brain regions have been identified as supporting voice and voice gender perception. First, the temporal voice-sensitive areas (TVA’s) have been demonstrated (Belin et al., 2000; Belin et al., 2002; Fecteau et al., 2004; Fecteau et al., 2005) and are localized bilaterally along the upper bank (middle and anterior) of the Superior Temporal Sulcus (STS) (Alho et al., 2006; Belin et al., 2000), with a predominant role of the right hemisphere (Bestelmeyer et al., 2011). In addition, the categorization of voice gender appears to depend on the right voice selective areas to encode acoustical dissimilarity (Charest et al., 2012). Second, the frontal cortex and in particular the bilateral inferior frontal regions, seem to be important in the encoding of perceived ambiguity and to carry out categorical perception (Charest et al., 2012). Finally, in addition of these cortical areas, subcortical thalamic and brainstem regions participate to voice information processing (Pernet et al., 2015b).

3.4. Voice and Speech Perception

Speech perception and voice perception co-exist in a single auditory signal, but there is a debate as to whether they are supported by independent, partially independent, or co-dependent cognitive processes and neuroanatomical pathways.

From a neuroanatomical perspective it appears that voice gender categorisation and phoneme categorisation are dissociated: left vs. right hemisphere dominance, and anterior/mid STS vs. lateral STG and posterior STS. Early work suggests a neuroanatomical dissociation between voice and speech, for example a case showing a patient with receptive aphasia following left hemisphere damage with essential intact speaker recognition (Assal et al., 1981a). Later work is supportive of these findings, for example Pertez et al. (1994) presented two patients with superior temporal cortex lesions (bilateral), both with poor voice perception and good speech perception.

In line with this neuro-functional dissociation, the classic hypothesis in speech perception is that talker (voice specific) information is extracted along with the speech signal first, and is then stripped away to access the phonemic content. Speaker normalisation is a proposed perceptual process in which phonological identical utterances with extensive acoustic variation across talkers can be easily recognised as instances of the same linguistic object (phonemes/ words) by listeners (Johnson and Mullennix, 1997). This proposed process is based on exemplar models where internal representations adapt according to the 'perceived speaker identity' (Johnson, 1990 p. 642). Multiple cues are used within representations involved in speaker normalisation including visual, familiar speaker recognition, prior expectations and acoustic cues such as formant transitions. Therefore when speaker normalisation occurs it is suggested that the auditory representation of the speech signal is modified before recognition occurs, and so variation in speech is stripped away (Goldinger, 1996; Johnson, 1990; Johnson et al., 1993). This view therefore suggests that voice and speech (as opposed to sound analysis) are processed separately. Recent neuroimaging studies have also suggested that linguistic and vocal information are relatively independent in terms of neural correlates (Belin et al., 2000; Belin et al., 2002; Belin and Zatorre, 2003; von Kriegstein et al., 2003; von Kriegstein and Giraud, 2004). Parallel models of language processing (Knosche et al., 2002; e.g. Mullennix and Pisoni, 1990) suggesting that phonemes (i.e. linguistic information for speech perception) are processed following or parallel to voice perception, have been further supported by ERP studies (Kaganovich et al., 2006).

Within the 'auditory face' model (Figure 3-3) Belin et al. (2004) suggest three main types of vocal information that are processed: vocal speech analysis, vocal affect analysis and voice identity analysis. The model suggests that firstly a general low-level analysis of auditory information occurs in sub-cortical nuclei and core regions of the auditory cortex. Voice structural analysis then occurs in which the three main types of vocal information are extracted and processed in functional pathways that are both interactive and dissociable. According to this model the analysis of speech information occurs in the anterior and posterior STS, the inferior prefrontal regions, and the pre-motor areas, predominantly in the left hemisphere. Vocal affective information occurs in the temporo-medial areas, the anterior insula, the amygdale and the inferior prefrontal regions, predominantly in the right hemisphere. Voice identity information is suggested to occur within regions of the right anterior STS.

In contrast to the hypothesis that voice and speech are processed separately, the effect of talker variability on speech perception (and therefore associations between voice and speech processing) has been demonstrated by many. For instance, using a continuous recognition memory procedure, Palmeri et al. (1993) showed that specific voice information is retained in the memory along with item information, and these attributes aid later recognition. Nygaard et al. (1994) showed that learning to identify a talker's voice has an effect on subsequent word recognition performance. Similarly, increased sensitivity to talker-specific information by learning affects the perception of the linguistic properties of speech in isolated words and sentences (Nygaard and Pisoni, 1996). Such results contradict the notion of complete independence and suggest that voice identity perception and speech perception are linked in their perceptual underpinnings. In particular, Remez et al. (1997) show that talker identity can be recognized from sine wave replicas of natural speech that preserved idiosyncratic phonetic variation, thus suggesting that phonetic properties serve to identify both speech and voices.

Kreitewolf et al. (2014) proposed a model which proposes neural interactions between the left and right hemispheres in terms of speech and voice recognition. This study used vocal tract parameters and glottal fold parameters (two main acoustic features that are involved in both voice and speech processing) and showed that functional interactions between left and right hemispheric homologous regions occurs in both sets of parameters i.e. interactions between the left and right hemispheric STS/STG process vocal tract parameters in both speech and voice perception; and interactions between left and right Heschl's gyri process glottal fold parameters in both speech and voice perception. The proposed schematic (Figure 3-5) assumes that speech perception is processing predominantly in the left hemisphere whereas voice perception lateralisation is less conclusive, as is the general consensus.

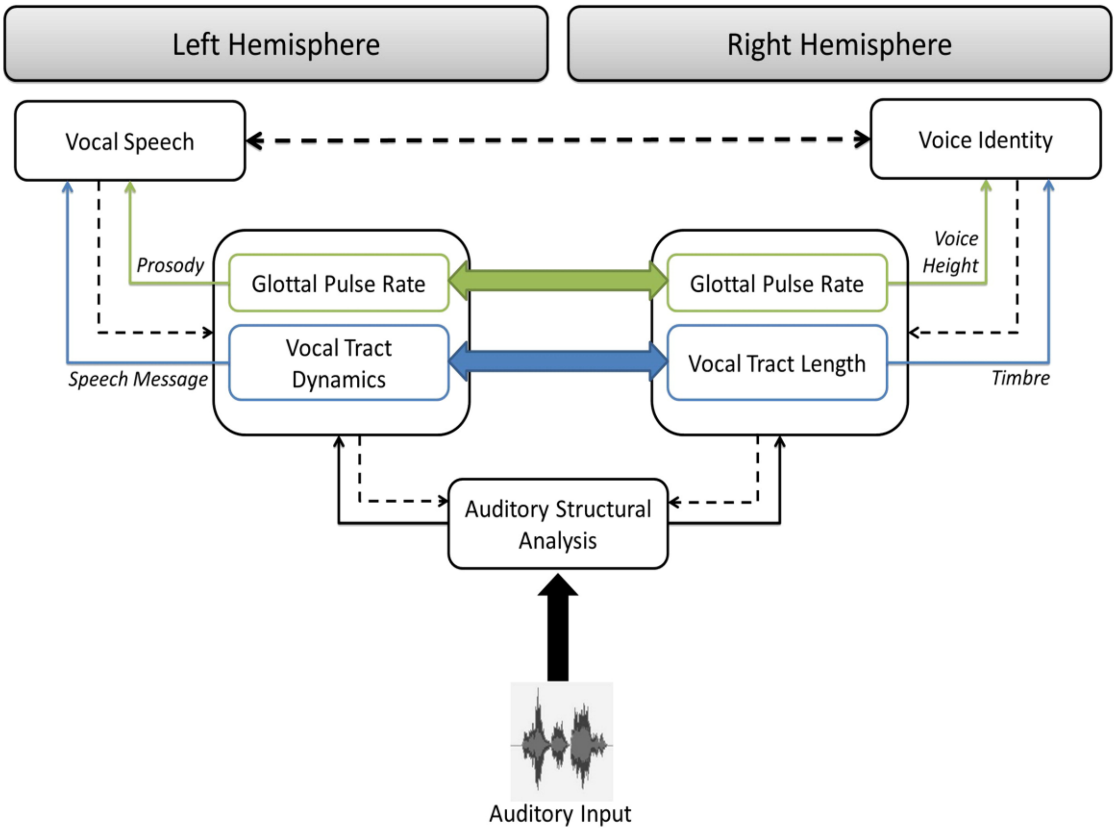

Figure 3-5: A schematic showing the potential mechanism in the processing of auditory voice and speech.

Taken from Kreitewolf et al. (2014). Speech (left hemisphere/ left side of figure) and voice (right hemisphere/ right side of figure) interact via neural processes to allow speech perception areas to be informed via voice perception areas. Glottal pulse rate, vocal tract dynamics and length, are speaker parameters which may be processed in distinct cortical regions (i.e. Heschl's gyrus processes glottal fold parameters, and STG/STS processes vocal tract parameters) and interact with homologous regions of the pathway in the opposite hemisphere.

Key: — forward connections --- backward and lateral connections

3.5. Study 1 - Voice gender categorisation vs. phoneme categorisation

Article:

Pernet, C. R., Belin, P., & Jones, A. (2013). *Behavioral evidence of a dissociation between voice gender categorization and phoneme categorization using auditory morphed stimuli*. Frontiers in psychology, 4: 1018. doi: 10.3389/fpsyg.2013.01018

Oral Presentation:

Jones, A. (2013). *Dissociating voice from speech.* Paper presented at SINAPSE Annual Scientific Conference; 2014 June 14; Edinburgh, UK.

Contributions by Anna Jones: study design, ethical submission, participant recruitment, data collection and analysis, results interpretation, writing of and commenting on manuscript, writing of and giving of oral presentation.

3.5.1. Introduction

The aim of this study was to investigate if the perceptual processes used to code voice information interact with the ones used to code phonemic information in the speech signal. In an attempt to dissociate between independent and linked perceptual underpinnings of voice and speech perception, the pattern of performance of listeners in two orthogonal identification tasks was investigated. Using sounds from different continua of 'male'-/pæ/ to 'female'-/tæ/ and 'male'-/tæ/ to 'female'-/pæ/, subjects categorised stimuli as being either 'male'-'female' (gender task) or /pæ/-/tæ/ (phoneme task)[†]. Although other studies have looked at either gender or identity in the context of speech (Bachorowski and Owren, 1999; Gelfer and Mikos, 2005; Lass et al., 1976; Whiteside, 1998) few have tested the two mechanisms simultaneously.

[†] /pæ/ and /tæ/ are phonetic transcriptions and are pronounced 'pa' and 'ta' respectively.

Since it has been suggested that voice and phoneme perception rely on similar phonemic properties, but that phoneme categorization must account for talker variability (talker normalization), the following was expected:

- voice gender information would be processed faster than phonemic information (Mullennix and Pisoni, 1990)
- similar phonetic cues would be used in both tasks.

Reaction time (RT) differences between tasks with identical weights of acoustic clues imply a sequence of information processing (i.e. non-dissociated processes) on the basis of shared acoustic information (i.e. the same representations). In contrast, RT differences with different weights of acoustic clues imply parallel and dissociated information processing on the basis of different representations.

To further investigate the role of acoustic versus phonemic features in each task, continua were also equalized in term of pitch height (f0) or timbre (supra-laryngeal filtering). If a normalisation process is taking place during the phoneme task, equating sounds in f0 or in some aspect of timbre should lead to faster RTs in those conditions.

Following the results of Pernet and Belin (2012) who investigated gender categorization in a similar context (but using a single syllable /had/), sigmoid response curves and super-Gaussian perceptual (d prime *or* d') curves (i.e. a distribution with a positive excess kurtosis – a more acute peak surrounding the mean with a heavy-tail) were expected in terms of accuracy results. Both curves are prototypical of categorical responses in 2 alternative forced choice (AFC) designs, although they do not necessarily reflect true categorical spaces (Gerrits and Schouten, 2004). Therefore it was hypothesized that in the gender task, significant differences among the original, timbre-equalized and f0-equalized sounds would be observed, with altered responses for the timbre-equalized sounds. For phonemes, no effect of pitch height or timbre was expected since it is known that phoneme perception in English relies on acoustic clues such as the voice-onset-time (VOT) and formant transitions (Koch et al., 1999).

Finally, to investigate which specific acoustic features were used by participants to carry out stimuli categorisation (deciding if the stimuli heard was ‘male’, ‘female’, ‘pa’ or ‘ta’), the features were extracted from each stimuli used along each continuum using the Praat software (Boersma and Weenink, 2009) then reverse correlation analyses were carried out to look for differences between the features for stimuli located above and below the PSE (point of subjective equality- the point at which the subject perceives the stimuli to be 50% male / 50% female). The following acoustic features were chosen for extraction and analysis (Table 3-1):

Acoustic Feature	Definition	Rationale
Fundamental frequency (mean f0)	The frequency with which the wave cycle of the sound repeats itself over time, perceived as vocal pitch	To investigate the role of the fundamental frequency (f0) of the vocal fold vibration, the signal-to-noise ratio of the stimuli, and the acoustic resonances of the vocal tract in gender and phoneme categorisation
Harmonic to Noise Ratio (HNR)	Estimate of the level of noise in human voice signals, perceived as voice quality	
Formants 1 - 4 (f1, f2, f3, f4)	Acoustic resonances of the vocal tract that are measured as peaks in amplitude on vocalic sound spectrum and determined by the shape of the vocal tract. Perceived as articulation of phonemes (especially vowels) & involved in timbre perception	

Table 3-1: Acoustic features extracted and analysed via reverse correlations.
This enabled investigation of specific acoustic features used by participants in gender and phoneme categorisation.

3.5.2. Method

Participants

Eighteen subjects participated in this study (9 females 35.3 ± 9.2 years old, 9 males 29.1 ± 3.6 years old). Subjects were all healthy volunteers with no known neurological or psychiatric disorder, no uncorrected visual impairment, no uncorrected hearing loss, no speech and language therapy history, no communication impairment and all had English as their first language. (See Appendix 1 and 2 for participant information sheets and consent forms).

Paradigm

Subjects were presented with two 2 AFC identification tasks: voice gender (male vs. female) and phoneme (/pæ/ vs. /tæ/). For each task, there were three conditions (all participants completed all 3 conditions for both tasks): original sounds, f0-equalized sounds and timbre-equalized sounds. Within each of the 3 conditions, for each task (gender and phoneme), there were two full continua of morphed sounds: the 1st continuum going from Male-/pæ/ to Female-/tæ/ and the 2nd continuum going from Male-/tæ/ to Female-/pæ/. An important point is that the same speakers were used for both continua (the same male pronouncing /pæ/ and /tæ/ and the same female pronouncing /pæ/ and /tæ/). In each of the 3 conditions (original, f0-equalized and timbre-equalized sounds) and for both tasks (voice gender and phoneme perception) each subject heard the following sounds (presented pseudo-randomly) six times each: 100% Male-/pæ/; 100% Male-/tæ/; 100% Female-/tæ/; 100% Female-/pæ/; 90% Male-/pæ/ and 10% Female-/tæ/; 90% Male-/tæ/ and 10% Female-/pæ/; 80% Male-/pæ/ and 20% Female-/tæ/; 80% Male-/pæ/ and 20% Female-/tæ/ etc. for 11 full steps on the morphed continua. Therefore each participant heard 132 stimuli (2 continua * 11 steps * 6 trials) for each condition they completed. This design allowed the investigation of the effect of the task (i.e. tell if for example the stimulus 80% Male-/pæ/ 20% Female-/tæ/ was male or female vs. /pæ/ or /tæ/) while controlling for the general acoustic characteristics of the stimuli since the same stimuli were used in both tasks. However, specific acoustic characteristics could still be identified as the stimuli grouping differed between tasks. In addition, pitch height equalization and timbre equalization (see Section 3.5.1) allowed the specific contribution of these features on the subject responses to be investigated. In total, 18 different continua of stimuli were generated from 6 different speakers (3 males and 3 females pronouncing /pæ/ and /tæ/) and 1 male and 1 female participant carried out all the tasks for each pair of continuum. Figure 3-6 shows an illustration of different aspects of one continuum of male-/pæ/ to female-/tæ/.

Participants carried out the 2 identification tasks (voice gender and phoneme perception) in 6 separate sessions (3 phoneme categorization sessions and 3 gender categorization sessions) with an interfering tone discrimination task lasting about 3 minutes (see Chapter 4 for more detail) in the middle of the 6 sessions. This task was primarily designed to minimise the influence of one task on the other. The order in which the tasks were presented was counterbalanced across subjects, along with the session order. Subjects listened to all the sounds via headphones and answered by pressing keys on a keyboard. Key orientation was counterbalanced between participants.

Instructions were as follows:

> "You will hear a series of sounds. You have to decide for each of these sounds whether it sounds more MALE (/pæ/) or more FEMALE (/tæ/). Here is an example of each of these two categories (the most extreme sounds from the continuum were played at this point). So if the sound you hear is closer to the MALE (/pæ/) sound, answer with the key 'A' ('L'); if the sound if closer to the FEMALE (/tæ /) sound answer with the key 'L' ('A'). Do you understand?"

If the subject did not understand, the examples were played once more and the final two sentences repeated. The full set of tasks was labelled the 'Voice and Speech Perception Battery' (VSPB).

Stimuli

Original stimuli were each recorded three times in a sound studio at the Voice Neurocognition Laboratory (http://vnl.psy.gla.ac.uk/). Three males and three females voiced the phonemes /tæ/ and /pæ/ and stimuli with the clearest separation between the consonant and the vowel (as seen on spectrograms) were selected. Stimuli were then manipulated using STRAIGHT (Kawahara, 2003; Kawahara, 2006; Kawahara and Akahane–Yamada, 2006) software, running under Matlab®. STRAIGHT performs an instantaneous pitch-adaptive spectral smoothing in each stimulus for separation of contributions to the voice signal arising from the glottal source vs. supra-laryngeal filtering. The algorithm decomposes a voice stimulus into five parameters: f0, frequency, time, spectro-temporal density and aperiodicity. Stimuli are then synthesized and each parameter can therefore be manipulated and combined across stimuli independently of one another. Here we used time-frequency landmarks to put in corresponding voices, allowing linear morphing. The morphing was based on 3 temporal (onset of the consonant, onset of the vowel, offset of the vowel) and 9 spectral (f0 identified on the consonant and onsets/offsets of the vowel's f0/f1/f2/f3 formants) anchoring points. The morphing was performed from male-/pæ/ stimuli to a female-/tæ/ stimuli and male-/tæ/ stimuli to female-/pæ/ stimuli, in 9 steps varying by 10% (plus the two original sounds re-synthesized, thereby creating continua containing 11 steps in total). By setting anchoring points on onsets of the consonant and vowels, offset of the vowel and on f0 on the consonant and the vowel, the algorithm could synthesize new stimuli for which the whole sounds were morphs representing a mixture of male-female and /pæ/-/tæ/. However, by also selecting f1/f2/f3 on the vowel, we forced the algorithm to match these particular spectral points on the vowel. In addition, since the source (f0) and the filter (supra-laryngeal filtering) are dissociated, additional morph continua that were equalized in f0 or in timbre across the stimuli were obtained. For the pitch and timbre equalised continua, the original /pæ/ and /tæ/ from male and female speakers were first equalized in f0 or timbre and then the morphs were created. Stimuli within each continuum were finally root mean squared normalized.

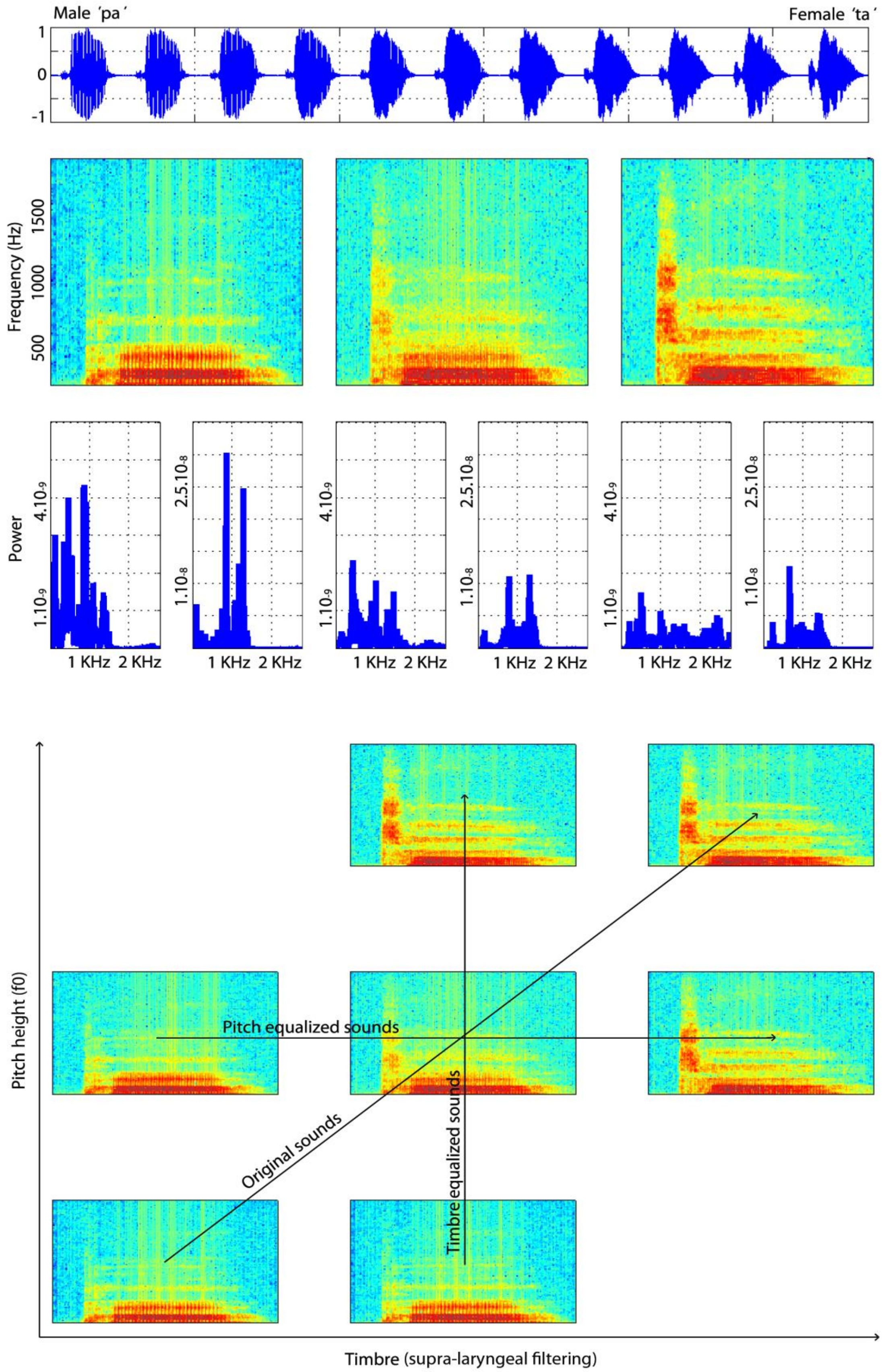

Figure 3-6: Illustration showing different aspects of one continuum of male-/pæ/ to female-/tæ/.

Row 1: male-/pæ/ to female-/tæ/ continuum of 11 steps in the time domain (waveforms).

Row 2: time-frequency domain (spectrograms with hamming window, sampling at 22040Hz) of the 100% male-/pæ/ stimulus, 50% male-/pæ/, 50% female-/tæ/ stimulus and 100% female-/tæ/ stimulus from this continuum are shown.

Row 3: Power spectra of the consonant and vowel separately for the same stimuli.

Bottom: Stimuli 'space' with spectrograms of the extreme stimuli (100% male and 100% female) and the 50% morphed stimulus for each of the 3 conditions: original sounds, f0-equalized & timbre equalized.

Data processing

For each subject, condition, continuum, and morphing step, the 6 scores and reaction times (RT) were collected and cleaned for outliers. S-outliers were detected from the RTs, and, if any were present, they were removed from the score and RT data (8.6% of the data removed). The S-outlier detection method i.e. those deviant from the absolute median distance among all pairs (Rousseeuw and Croux, 1993), has a high sensitivity. The mean score (percentage female/'ta') and mean RT were then computed for each subject. Such data pre-processing is routinely performed by many and is indeed recommended to remove observations that are inconsistent with the remainder of the data (Barnett and Lewis, 1994).

The procedure was iteratively repeated for each stimulus (i.e. 18 subjects, 3 conditions, 2 morphs, 11 steps). From the mean percentages of female/'ta' responses per continuum, a cumulative Weilbull function (flexible measurement detailing the probable cumulative distribution) was fitted in Matlab® using unconstrained nonlinear minimization. The point of subjective equality (PSE: the point at which the subject perceives the stimuli to be 50% male / 50% female) was computed. Percentages of correct responses that could not be modelled and/or gave aberrant PSE values were discarded (in total 17.59% of the data). On average, the same amount of data was discarded in each task (13.88% in the gender task vs. 21.29% in the phoneme task, percentile bootstrap confidence interval of the difference [-5.4 3.01] showing no significant difference). At this stage, the 2 continua (1. male-/pæ/ to female-/tæ/; 2. male-/tæ/ to female-/pæ/) did not differ significantly in terms of percentages or RTs when computed per condition/step (percentile bootstrap on the mean difference with adjustment for multiple comparisons). Averages were thus computed for each condition/step and all following statistical analyses were performed on these averaged scores and RTs cleaned for outlying data point.

3.5.3. Data Analysis

Data cleaning

For all analyses apart from the reverse correlation, 20% trimmed means were used (i.e. computing the mean over 12 participants and removing the 3 highest and 3 lowest values). Trimming simply removes the lowest and highest values, and the p-value comes from estimating the null hypothesis via bootstrapping. Importantly, using trimmed means gives identical results as when using standard means if data are normally distributed, and therefore results can be interpreted the same way as with means (Figure 3-7). Data is, however, almost never normally distributed (e.g. Micceri, 1989) and standard statistics then seriously lack power. In these cases the mean is a poor estimator of the population average and trimmed means have been shown to better reflect the true underlying average. In addition, because significance is obtained using bootstrap procedures, analyses are assumption free. Here trimmed means ensured the data were not biased by inaccurate/slow participants or extremely accurate/fast participants.

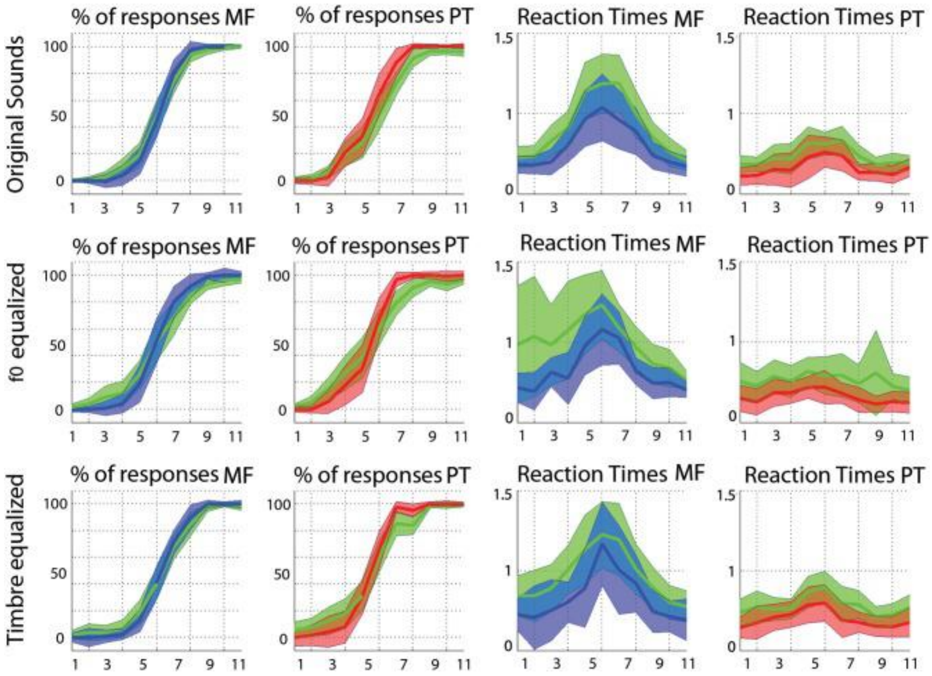

Figure 3-7: Comparisons between the ‘standard’ mean computed on raw data with the 95% confidence interval and the robust alternative.

‘Standard’ mean (in green) is compared to trimmed means on cleaned data with bootstrapped 95% confidence intervals (in blue for the gender task, and in red for phoneme task). All graphs: x-axis show steps along continuum; % response graphs - y-axis shows % of responses as either ‘female’ or ‘ta’; reaction time graphs – y-axis shows reaction times in msec.

As shown in Figure 3-7, mean scores and RTs on raw data and trimmed mean scores and RTs on ‘cleaned’ data were quite similar, indicating that data were close to normally distributed. However, it is also apparent that mean response curves are flatter than trimmed mean ones and mean RTs are all slower than trimmed mean ones. Using the ‘standard’ mean with raw data also led to larger confidence intervals, sometimes due to a single data point in one subject, which illustrates why using trimmed means offers a more reliable alternative than means. Importantly, as will be discussed in the results section, the relationships between conditions (e.g. faster RTs for phonemes than voices and no effect of f0 and timbre equalization) is unchanged between the methodology types.

Within and between task analyses

Statistical testing within tasks (i.e. original vs. same pitch vs. same timbre) and between tasks (i.e. gender vs. phoneme for each condition) was performed using pair-wise comparisons on the 20% trimmed mean difference. For a given comparison, the difference between pairs was computed and 1000 bootstraps obtained (sampling with replacement). The 20% trimmed means were then computed and the percentile CI and p-value obtained. Under the null hypothesis of 'no difference between 2 conditions', these differences are equally distributed around zero. The p-value for an observed difference therefore corresponds to the average number of times the bootstrapped trimmed means were above 0 (or 1 minus this average). It is thus possible to obtain a p-value of 0 if all the values are above or below 0. Finally, when multiple pair-wise comparisons were used (e.g. 9 comparisons testing within and between task differences, or 11 comparisons testing between tasks differences along the 11 steps of a continuum), an adjustment for multiple comparisons was applied (Pernet et al., 2012).

Analysis of percentages of responses

PSEs obtained for each pair of continua were averaged and a percentile bootstrap on trimmed means was computed, testing if the abscissa (i.e. centre of PSE along x-axis when plotted) of the PSE of each condition differed from 6, i.e. the physical middle of the continua. PSEs were also compared with each other (within and between tasks) using pair-wise comparisons.

In addition to this global measure of deviation from the physical middle of the continua, percentages of responses were compared between tasks (gender vs. phoneme) for each of the 11 steps in each condition separately (original sounds, pitch equalized, timbre equalized).

Finally, the rate of change between successive pairs was also tested against 0 (percentile bootstrap on trimmed means) and between tasks. D prime (or d') is a commonly used statistical measure used within signal-detection theory (Macmillan and Creelman, 2004). It provides you with the separation between the means of the signal and the noise distributions, compared to the standard deviation of the noise distribution. Importantly, a higher d' score indicates that the signal was more readily detected, whilst a d' score of 0 indicates no discrimination. In this study the perceptual distance (i.e. the d') is computed between each successive step (i.e. on each pair) and so shows the rate of change in the stimuli, i.e. the d' score at stimulus pair 8/9 represents the discrimination of stimulus 8 and 9 and therefore the rate of change in this stimulus can be detected. Therefore, in this case a d' score of 0 indicates no change was detected within the stimulus, whilst a high d' score indicates a high rate of change. While average percentages allow the absolute differences in categorization performance to be investigated, using the d' values allows the investigation of perceived distances along the continua. Importantly, using the d' measure also takes into account listener's response bias.

Analysis of Reaction times

For each condition and task, the average RTs were computed and pair-wise comparisons were performed within and between tasks. RTs were compared between tasks (gender vs. phoneme) for each of the 11 steps in each condition separately (original sounds, pitch equalized, timbre equalized). The rate of change (1^{st} derivative i.e. increase or decrease, and by how much) between successive pairs was also tested between tasks. The rate of change was computed as the average of absolute differences between successive pairs in each continuum. While averaged RTs allowed for the investigation of differences in processing time, using the 1^{st} derivative allowed for the investigation of any significant variations along the continua.

Reverse correlations

An analysis of which acoustic features were used by participants to categorize stimuli as 'male'-'female' or /pæ/-/tæ/ was also conducted.

Within the gender task, stimuli located below the PSE were categorized as male whilst stimuli located above the PSE were categorized as female. Both categories, however, included /pæ/ and /tæ/ phonemes and, across participants, different PSE values were obtained and different voice stimuli were used. By comparing the average acoustic properties of perceived male vs. perceived female sounds across participants, it was possible to reveal which acoustic features distinguished male stimuli from female stimuli.

Similarly, within the phoneme task, stimuli located below the PSE were categorized as /pæ/ whilst stimuli located above the PSE were categorized as /tæ/. Both categories, however, included male and female voices and, across participants, different PSE values were obtained and different voice stimuli were used. By comparing the average acoustic properties of perceived /pæ/ vs. perceived /tæ/ sounds across participants, it was possible to reveal which acoustic features distinguished the two phonemes.

Finally, since the same stimuli were used for both tasks, differences in which acoustic properties distinguished male-female from /pæ/-/tæ/ stimuli should reveal which features were diagnostic to the task at hand.

Using the Praat software (Boersma and Weenink, 2009), the fundamental frequency (mean f0) and Harmonic to Noise Ratio (HNR) of each sound was obtained. In addition, the consonants (/p/ or /t/) and vowel (/æ/) were analysed separately to obtain the mean f1, f2, f3 and f4 formant positions. For each stimulus used within the tasks, the consonant and the vowel were extracted manually (from 0ms to the onset of the vowel (=consonant) and from the onset of the vowel to the end of the stimulus (=vowel)) and formant values obtained automatically with Praat (search settings: 0 to 5500Hz, with 25ms windows and 30dB dynamic range). The reverse correlation analyses consisted of testing for differences in these sound properties (f0, HNR and formant dispersion [the average difference in frequency between successive formants, f2/f1, f3/f2 and f4/f3]) for stimuli located above or below the PSE in two ways:

1. For each subject, each condition and each continuum, the average f0, HNR and formant dispersions were computed separately for stimuli below and above the PSE.
2. A percentile bootstrap on the median differences (Pernet et al., 2012) across subjects was computed, thus revealing the acoustic properties that differed among stimuli

classified as ‘male’ vs. ‘female’ or /pæ/ vs. /tæ/. The median was used (rather than the trimmed mean as previously used) because the differences in acoustic features were often close to uniformly distributed (i.e. equal probability of variables along distribution – see Figure 3-10).

In addition to these comparisons, the average f0, HNR and formant dispersions were obtained for an ideal listener. This ideal listener separated all sounds equally, based on acoustical distances (the PSE was always 6, the acoustic middle of the continua, for all stimuli). Therefore, for the ideal listener acoustic properties were always averaged and compared for stimuli 1 to 5 vs. 7 to 11. In contrast, the PSE differed from participant to participant between 5 and 7, and acoustic properties could be averaged for stimuli 1 to 4 vs. 6 to 11 in one participant and for stimuli 1 to 6 vs. 8 to 11 in another participant. Comparing the results from the ideal listener to the ones observed in the participant population thus revealed biases in the information used, only if there was a consistent behaviour across subjects. The difference between the observed differences in acoustic properties and the differences obtained with the ideal listener were compared using, once again, a percentile bootstrap on the median differences.

3.5.4. Results

Results are reported here from the 3 main sets of analyses performed: 1) percentages of responses; 2) reaction times and 3) reverse correlations.

Percentages of responses

The average PSE was located at the middle of the physical continuum (i.e. step 6) in the gender task, for all three conditions (original, pitch equalized, and timbre equalized sounds). In the phoneme task, the abscissa was significantly smaller than 6 (biased toward /pæ/) for pitch and timbre equalized stimuli (Table 3-2). Pair-wise comparisons did not show significant differences within tasks (i.e. among conditions) but a significant difference between tasks was observed for the timbre-equalized condition (as seen in Table 3-2 & Figure 3-8 [column 2, rows 1-3]).

	Original sounds	*Pitch equalized*	***Timbre equalized***
Gender task	6.26 [5.9 6.6] p=0.09	6.03 [4.9 7.14] p=0.47	6.31 [5.8 6.83] p=0.04
Phoneme task	5.34 [4.6 6.08] p=0.01	**5.27 [4.24 6.31] p=0.003**	**5 [4.14 5.85] p=0**
Difference	0.7 [-0.19 1.7] p=0.04	0.3 [-1.6 1.8] p=0.5	**1.2 [0.008 2.75] p=0.004**

Table 3-2: Trimmed mean PSE with 95% CI (in square brackets) for each task and condition along with the p-value associated to the test of difference from 6.
The bottom of the table shows the trimmed mean differences between tasks. Significant p values are marked in bold (alpha adjusted for multiple comparisons).

Analyses of percentages of responses for each step separately showed higher ratings in the phoneme task than the gender task for steps 1, 4, 8 and 10 in the original sounds condition, for step 8 in the f0-equalized condition, and steps 1, 6, 7, 9, 10 and 11 in the timbre equalized condition (data shown in bold in Table 3-3).

Analysis of the rate of change between successive stimuli revealed, as expected, a significant increase in the perceptual distance for ambiguous stimuli (Figure 3-8 [column 3, rows 1-3]: d' significantly differs from 0, shown by the fact that the 95% CI does not cross 0). In the gender task, stimulus pairs 5/6, 6/7 and 7/8 differed from 0 for the original sounds, stimulus pairs 4/5, 5/6 and 6/7 differed from 0 for the f0-equalized sounds, and stimulus pairs 4/5, 5/6, 6/7 and 7/8 differed from 0 for the timbre equalized sounds. In the phoneme task, stimulus pairs 5/6, 6/7 and 7/8 differed from 0 for the original sounds, stimulus pairs 5/6 and 6/7 differed from 0 for the f0-equalized sounds, and stimulus pairs 4/5, 5/6, and 6/7 differed from 0 for the timbre equalized sounds.

Despite these variations, no significant differences (except pair 8/9 for f0-equalized stimuli and pair 3/4 for the original stimuli) between tasks were observed on d' when testing along the 10 distances, i.e. perceptual distances between consecutive stimuli were equivalent between tasks, leading to similar total d prime (i.e. the cumulative distance from step 1 to 11 (whole continuum) – see Table 3-4 & Figure 3-8 [column 4, rows 1-3]).

Task	*Condition*	*Steps along continua*										
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Original sounds	**MF**	**0 [03]**	**0 [07]**	**0 [06]**	**4 [06]**	**15 [421]**	**45 [2853]**	**80 [6280]**	**98 [8199]**	**100 [97100]**	**100 [98100]**	100 [97100]
	PT	**0 [06]**	**0 [09]**	**2 [016]**	**21 [020]**	**33 [1744]**	**64 [5279]**	**88 [93100]**	**100 [90100]**	**100 [98100]**	**100 [98100]**	100 [99100]
	Diff	*[-10] p=0*	*[-20] p=.07*	*[-81] p=.08*	*[-24-2] p=0*	*[-347] p=.2*	*[-40-11] p=.1*	*[-33-12] p=.1*	*[-174] p=.04*	*[-20] p=.1*	*[00] p=.006*	*[-30] p=.3*
Pitch equalized	**MF**	**0 [02]**	**0 [02]**	**1 [06]**	**4 [011]**	**19 [533]**	**51 [3468]**	**80 [6992]**	**92 [83100]**	**99 [96100]**	**100 [95100]**	100 [97100]
	PT	**0 [02]**	**0 [03]**	**6 [015]**	**18 [332]**	**30 [1348]**	**68 [5283]**	**97 [91100]**	**100 [98100]**	**100 [97100]**	**99 [95100]**	100 [97100]
	Diff	*[-22] p=.2*	*[02] p=.2*	*[-154] p=.3*	*[-289] p=.3*	*[-2212] p=.8*	*[-3014] p=.5*	*[-280] p=.06*	*[-17-2] p=.004*	*[-31] p=.2*	*[05] p=.7*	*[-33] p=.6*
Timbre equalized	**MF**	**0 [03]**	**0 [07]**	**1 [06]**	**2 [06]**	**13 [421]**	**41 [2853]**	**71 [6280]**	**90 [8199]**	**100 [97100]**	**100 [98100]**	100 [97100]
	PT	**0 [06]**	**2 [09]**	**4 [016]**	**8 [020]**	**31 [1744]**	**66 [5279]**	**98 [93100]**	**95 [90100]**	**100 [98100]**	**100 [98100]**	100 [99100]
	Diff	***[-30] p=.01***	*[-81] p=.06*	*[-131] p=.1*	*[-130] p=.05*	*[-340] p=.06*	***[-40-10] p=.004***	***[-33-13] p=0***	*[-165] p=.3*	***[-20] p=.01***	***[00] p=.01***	***[-30] p=.01***

Table 3-3: Trimmed mean percentages and bounded 95% CI of 'female' or 'ta' responses for each task and condition, and 95% CI and p values of differences between tasks.

Difference between conditions in italics. Significant p-values are marked in bold (alpha adjusted for multiple comparisons).

Task	*Condition*	*Successive stimuli along continua*									
		1/2	**2/3**	**3/4**	**4/5**	**5/6**	**6/7**	**7/8**	**8/9**	**9/10**	10/11
Original sounds	**MF**	**0.01 [-0.04 0.07]**	**0.02 [-0.09 0.1]**	**0.06 [-0.1 0.2]**	**0.17 [-0.07 0.4]**	**0.79 [0.5 1.05]**	**0.76 [0.4 1.1]**	**0.4 [0.1 0.6]**	**0.07 [-0.08 0.2]**	**0.007 [-0.05 0.07]**	0.005 [-0.05 0.06]
	PT	**0.06 [-0.01 0.1]**	**0.02 [-0.1 0.2]**	**0.31 [0.06 0.5]**	**0.16 [-0.1 0.5]**	**0.67 [0.2 1.07]**	**0.48 [0.06 0.9]**	**0.29 [0.01 0.5]**	**0.02 [-0.03 0.08]**	**-0.01 [-0.06 0.03]**	-0.01 [-0.07 0.03]
	Diff	*[-0.1 0.07] p=.5*	*[-0.3 0.1] p=.5*	*[-0.4 0.003] p=.01*	*[-0.5 0.4] p=.9*	*[-0.5 0.6] p=.7*	*[-0.6 0.9] p=.3*	*[-0.4 0.5] p=.5*	*[-0.06 0.2] p=.5*	*[-0.08 0.1] p=.3*	*[-0.06 0.1] p=.4*
Pitch equalized	**MF**	**-0.001 [-0.08 0.07]**	**0.04 [-0.07 0.1]**	**0.05 [-0.1 0.2]**	**0.3 [0.08 0.6]**	**0.8 [0.5 1]**	**0.5 [0.09 0.9]**	**0.2 [-0.06 0.5]**	**0.1 [-0.1 0.3]**	**0.02 [-0.1 0.1]**	-0.02 [-0.1 0.07]
	PT	**0.02 [-0.06 0.1]**	**0.14 [-0.1 0.3]**	**0.2 [-0.1 0.5]**	**0.2 [-0.1 0.6]**	**0.5 [0.07 1]**	**0.6 [0.2 1]**	**0.1 [-0.03 0.3]**	**-0.03 [-0.1 0.04]**	**-0.02 [-0.1 0.08]**	0.02 [-0.04 0.09]
	Diff	*[-0.1 0.1] p=.6*	*[-0.7 0.2] p=.4*	*[-0.8 0.4] p=.8*	*[-0.5 0.7] p=.2*	*[-0.8 0.6] p=.7*	*[-0.7 0.5] p=.5*	*[-0.4 0.4] p=.8*	***[0.01 0.5] p=.002***	*[-0.2 0.2] p=.5*	*[-0.1 0.09] p=.3*
Timbre equalized	**MF**	**0.02 [-0.1 0.2]**	**-0.003 [-0.09 0.09]**	**0.03 [-0.08 0.1]**	**0.2 [0.06 0.4]**	**0.6 [0.3 0.8]**	**0.9 [0.5 1.2]**	**0.4 [0.08 0.7]**	**0.1 [-0.06 0.4]**	**0.05 [-0.001 0.1]**	-0.02 [-0.1 0.05]
	PT	**0.02 [-0.1 0.1]**	**0.07 [-0.1 0.2]**	**0.08 [-0.1 0.2]**	**0.3 [0.06 0.5]**	**0.8 [0.4 1.2]**	**0.7 [0.4 1.1]**	**-0.0005 [-0.1 0.1]**	**0.09 [-0.02 0.2]**	**0.003 [-0.07 0.08]**	0.01 [-0.05 0.07]
	Diff	*[-0.3 0.3] p=.8*	*[-0.4 0.1] p=.2*	*[-0.2 0.2] p=.3*	*[-0.6 0.2] p=.4*	*[-0.9 0.3] p=.2*	*[-0.6 0.6] p=.6*	*[-0.09 0.9] p=.02*	*[-0.1 0.4] p=.6*	*[-0.06 0.1] p=.1*	*[-0.2 0.09] p=.5*

Table 3-4: Trimmed mean d' values and 95% CI for each task and condition, and 95% confidence intervals and p-values of differences between tasks.

Differences in italics. Significant p-values are marked in bold (alpha adjusted for multiple comparisons).

P-values of differences between tasks shown in italics.

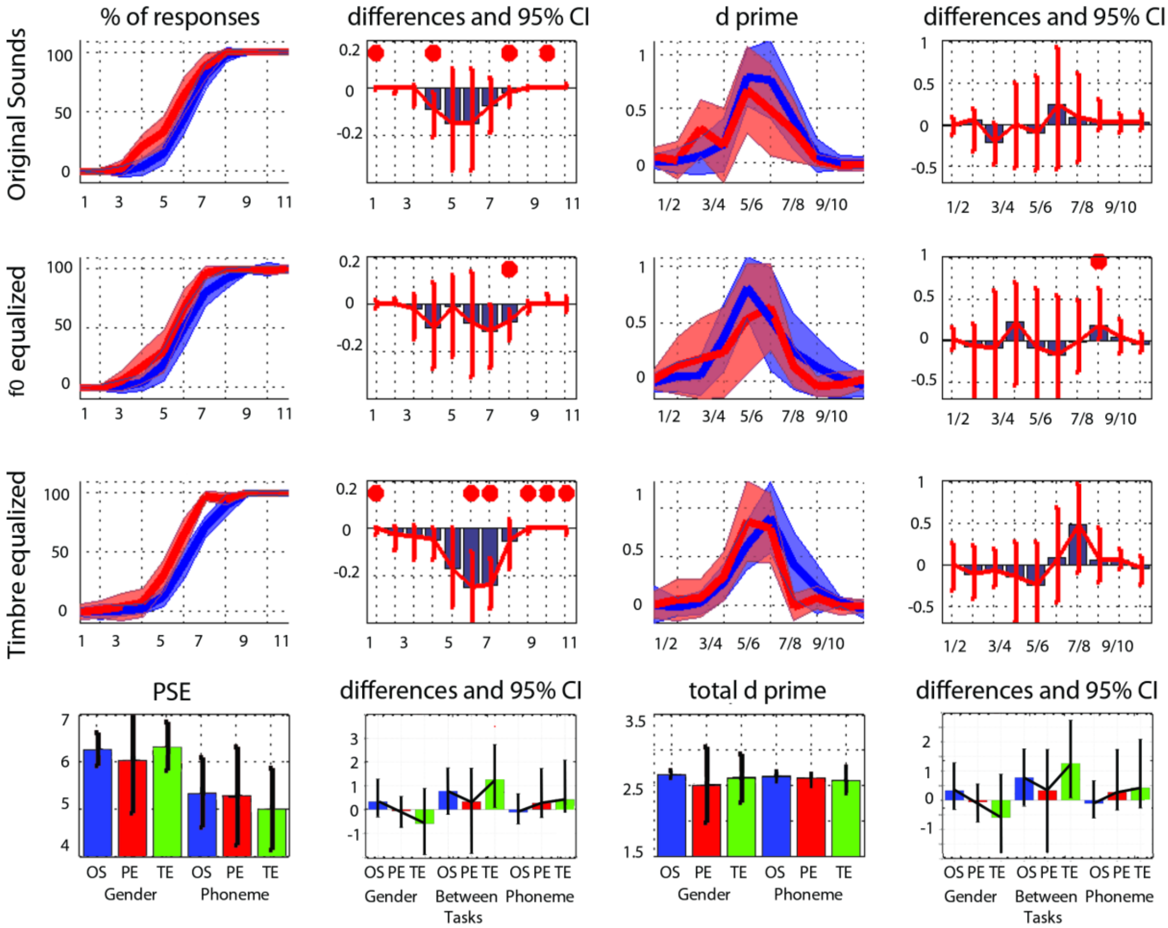

Figure 3-8: Summary of percentage of responses results showing the trimmed mean responses between conditions and tasks.

Rows 1-3: (left to right) - (i) the 95% CI of response curves in the gender task (blue, percentage of female responses) and in the phoneme task (red, percentage of /tæ/ responses), (ii) the differences between tasks in percentage of responses computed for each step (significant differences with adjustment for multiple comparisons marked by a red dot), (iii) the 95% CI of the d' measured between each successive pairs and (iv) the differences between tasks in d' computed for each step.

Row 4: Summary measures (left to right) - (i) the average PSE per task and condition, (ii) the within and between tasks differences in PSE, (iii) the total (cumulative) d', (iv) the within and between tasks differences in total d'.

Key: OS: original sounds, PE: f0-equalized sounds, TE: timbre equalized sounds. Bars represent the adjusted 95% CI on all graphs.

Reaction times

Averaged over the 11 steps of each continua, RTs were significantly shorter in the phoneme task than the gender task in each condition (-128ms [-42ms -246ms] p=0 for original sounds; -180ms [-66ms -269ms] p=0 for f0-equalized sounds; -172ms [-81ms -381ms] p=0 for timbre equalized sounds), and no differences were observed within tasks (i.e. between the original sounds, f0-equalized and timbre-equalized conditions). When testing for differences between tasks for each condition along the 11 steps, RTs were found to be significantly shorter in the phoneme task (Figure 3-9 [column 2, rows 1-3]) from steps 3 to 10 with the original sounds (max. -241ms at step 6), for all 11 steps in the f0-equalised condition (max -337ms at step 7), and for steps 1, 5, 6, 7, 8 and 9 in the timbre-equalised condition (max -464ms at step 6) as shown in Table 3-5 (data shown in bold).

The average rate of change along the 11 steps showed significantly larger changes in the gender task than in the phoneme task (0.17ms vs. 0.10ms, difference = [0.05ms 0.22ms], p=0 for the original sounds; 0.18ms vs. 0.10ms, difference = [0.07ms 0.27ms], p=0 for the f0-equalized sounds; 0.17ms vs. 0.13ms, difference = [0.07ms 0.37ms], p=0.002 for the timbre-equalized sounds) versus, again, no differences within tasks. Analysis of the rate of change between steps revealed significantly larger changes in the gender task from steps 5 to 6 and steps 6 to 7 and significantly smaller changes from steps 10 to 11 with the original sounds; significantly larger changes from steps 6 to 7 and from steps 9 to 10 with the f0-equalized sounds; and significant larger changes from steps 7 to 8 and from steps 8 to 9 with the timbre-equalized sounds (Table 3-6 [data shown in bold], Figure 3-9 [column 4, rows 1-3]).

Task	*Condition*	*Steps along continua*										
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	11
Original	**MF**	**0.68 [0.63 0.73]**	**0.68 [0.62 0.73]**	**0.7 [0.62 0.78]**	**0.8 [0.69 0.91]**	**0.96 [0.79 1.1]**	**1 [0.83 1.2]**	**0.96 [0.78 1.1]**	**0.89 [0.74 1]**	**0.74 [0.65 0.82]**	**0.7 [0.63 0.76]**	0.67 [0.61 0.73]
	PT	**0.61 [0.55 0.66]**	**0.61 [0.55 0.67]**	**0.65 [0.55 0.75]**	**0.64 [0.54 0.75]**	**0.73 [0.59 0.86]**	**0.76 [0.66 0.85]**	**0.74 [0.64 0.83]**	**0.63 [0.58 0.68]**	**0.63 [0.58 0.69]**	**0.62 [0.56 0.67]**	0.66 [0.61 0.72]
	Diff	***[0.002 0.1] P=0.05***	***[-0.01 0.1] p=0.1***	***[0.0010.1] p=0.04***	***[0.070.2] p=0.0***	***[0.080.3] p=0.0***	***[0.10.4] p=0.0***	***[0.070.4] p=0.0***	***[0.110.3] p=0.0***	***[0.030.1] p=.0.006***	***[0.030.1] P=0.0***	*[-0.05 0.09] p=0.5*
Pitch equalized	**MF**	**0.71 [0.62 0.8]**	**0.69 [0.57 0.8]**	**0.81 [0.72 0.9]**	**0.77 [0.61 0.9]**	**0.97 [0.79 1.1]**	**1.07 [0.86 1.2]**	**1.02 [0.86 1.1]**	**0.82 [0.73 0.9]**	**0.74 [0.64 0.8]**	**0.75 [0.66 0.8]**	0.7 [0.65 0.75]
	PT	**0.6529 [0.57 0.73]**	**0.6303 [0.54 0.71]**	**0.6878 [0.6 0.77]**	**0.6873 [0.62 0.75]**	**0.7199 [0.65 0.78]**	**0.7221 [0.61 0.82]**	**0.6836 [0.62 0.74]**	**0.6435 [0.56 0.72]**	**0.6171 [0.56 0.66]**	**0.6322 [0.56 0.69]**	0.6255 [0.56 0.68]
	Diff	***[0.01 0.17] p=0.03***	***[0.030.16] p=0.004***	***[0.070.2] p=0.0***	***[0.030.26] p=0.002***	***[0.070.47] p=0.0***	***[0.170.59] p=0.0***	***[0.190.52] p=0.0***	***[0.060.21] p=0.0***	***[0.040.23] p=0.0***	***[0.030.22] p=0.01***	***[0.020.11] p=0.01***
Timbre	**MF**	**0.72 [0.62 0.82]**	**0.7 [0.51 0.9]**	**0.75 [0.56 0.94]**	**0.8 [0.67 0.93]**	**0.89 [0.64 1.1]**	**1.16 [0.90 1.4]**	**1 [0.72 1.2]**	**0.92 [0.74 1.1]**	**0.74 [0.61 0.87]**	**0.71 [0.61 0.81]**	0.69 [0.56 0.82]
	PT	**0.64 [0.58 0.71]**	**0.67 [0.57 0.77]**	**0.71 [0.62 0.79]**	**0.72 [0.64 0.8]**	**0.78 [0.67 0.89]**	**0.79 [0.68 0.91]**	**0.69 [0.58 0.79]**	**0.67 [0.61 0.74]**	**0.65 [0.59 0.71]**	**0.65 [0.58 0.71]**	0.67 [0.58 0.76]
	Diff	***[0.02 0.21] p=0.02***	*[-0.001 0.1] p=0.05*	*[-0.02 0.2] p=0.1*	*[-0.02 0.19] p=0.1*	***[0.0050.3] p=0.03***	***[0.10.6] p=0.0***	***[0.10.42] p=0.0***	***[0.10.44] p=0.002***	***[0.030.2] p=0.004***	*[-0.002 0.1] p=0.06*	*[-0.02 0.15] p=0.1*

Table 3-5: Trimmed mean RTs and 95% CI for each task and condition, with 95% confidence intervals and p-values of differences between tasks.

Differences in italics. Significant p-values are marked in bold (alpha adjusted for multiple comparisons).

Task	*Condition*	*Successive stimuli along continua*									
		1/2	**2/3**	**3/4**	**4/5**	**5/6**	**6/7**	**7/8**	**8/9**	**9/10**	10/11
Original sounds	**MF**	**0.07 [0.04 0.1]**	**0.08 [0.04 0.12]**	**0.16 [0.11 0.22]**	**0.17 [0.06 0.29]**	**0.26 [0.14 0.37]**	**0.22 [0.13 0.31]**	**0.22 [0.12 0.31]**	**0.15 [0.07 0.24]**	**0.11 [0.06 0.16]**	0.05 [0.02 0.08]
	PT	**0.06 [0.03 0.09]**	**0.06 [0.01 0.11]**	**0.08 [-0.01 0.17]**	**0.13 [0.02 0.24]**	**0.11 [0.04 0.17]**	**0.11 [0.03 0.18]**	**0.1 [0.01 0.19]**	**0.07 [0.03 0.11]**	**0.08 [0.05 0.11]**	0.1 [0.07 0.13]
	Diff	***[-0.05 0.05] p=.4***	***[-0.04 0.09] p=.2***	***[-0.04 0.14] p=.05***	***[-0.04 0.15] p=.05***	***[0.03 0.25] p=0***	***[00.2] p=.004***	***[-0.03 0.24] p=.03***	***[-0.011 0.16] p=.01***	***[-0.04 0.1] p=.2***	***[-0.08 -0.004] p=.002***
Pitch equalized	**MF**	**0.12 [0.06 0.18]**	**0.16 [0.09 0.23]**	**0.12 [0.02 0.21]**	**0.23 [0.07 0.38]**	**0.23 [0.04 0.42]**	**0.24 [0.15 0.32]**	**0.24 [0.11 0.37]**	**0.14 [0.08 0.19]**	**0.12 [0.07 0.17]**	0.06 [0.01 0.12]
	PT	**0.05 [0.02 0.09]**	**0.08 [0.05 0.12]**	**0.11 [0.05 0.17]**	**0.1 [0.06 0.15]**	**0.12 [0.06 0.17]**	**0.13 [0.06 0.19]**	**0.13 [0.07 0.19]**	**0.07 [0.01 0.13]**	**0.04 [0.01 0.07]**	0.08 [0.05 0.11]
	Diff	***[-0.01 0.1] p=.01***	***[-0.003 0.1] p=.01***	***[-0.07 0.1] p=.5***	***[-0.04 0.3] p=.09***	***[-0.007 0.3] p=.008***	***[0.010.1] p=0***	***[-0.06 0.2] p=.07***	***[-0.07 0.1] p=.3***	***[0.0050.1] p=.004***	*[-0.05 0.05] p=.9*
Timbre equalized	**MF**	**0.06 [-0.05 0.19]**	**0.06 [0.03 0.09]**	**0.11 [0.03 0.19]**	**0.14 [0.01 0.27]**	**0.24 [0.09 0.38]**	**0.3 [0.11 0.49]**	**0.24 [0.09 0.39]**	**0.17 [0.05 0.29]**	**0.08 [0.04 0.13]**	0.08 [0.01 0.15]
	PT	**0.05 [0.007 0.10]**	**0.09 [0.04 0.15]**	**0.12 [0.05 0.19]**	**0.14 [0.06 0.23]**	**0.16 [0.1 0.23]**	**0.19 [0.09 0.29]**	**0.08 [-0.001 0.17]**	**0.06 [0.03 0.09]**	**0.05 [0.03 0.07]**	0.06 [0.01 0.12]
	Diff	*[-0.04 0.09] p=.4*	*[-0.14 0.03] p=.1*	*[-0.1 0.07] p=.7*	*[-0.09 0.1] p=.6*	*[-0.09 0.3] p=.4*	*[-0.04 0.4] p=.1*	***[0.050.3] p=0***	***[0.0060.3] p=.004***	*[-0.02 0.08] p=.1*	*[-0.04 0.1] p=.1*

Table 3-6: Trimmed mean of the rate of change in RTs (1st derivative) and 95% CI for each task and condition, with 95% confidence intervals and p-values of differences between tasks.
Significant p values are marked in bold (alpha adjusted for multiple comparisons).

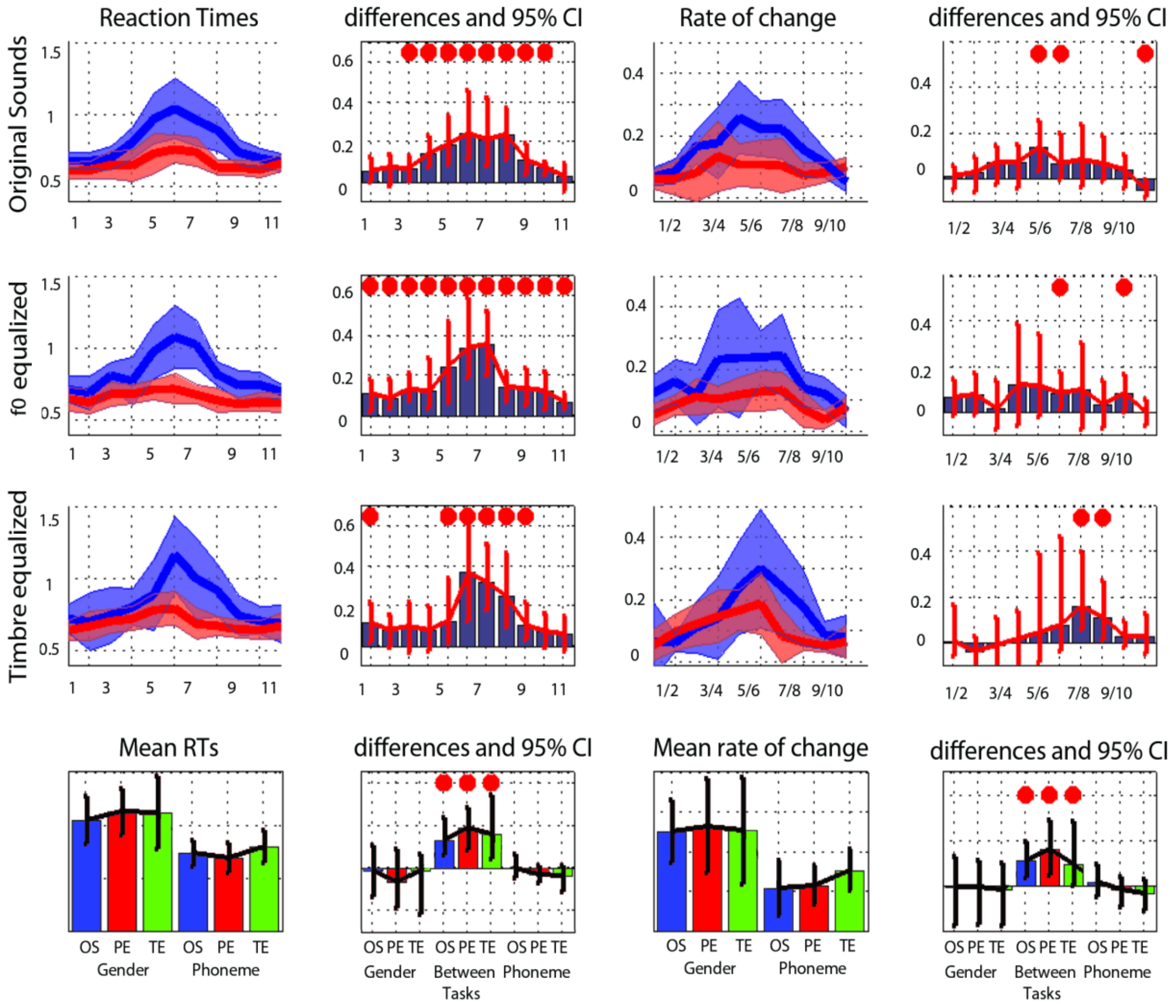

Figure 3-9: Summary of reaction time results showing the trimmed mean reaction times between conditions and tasks.

Rows 1-3: (left to right) - (i) the 95% CI of reaction times in the gender task (blue) and in the phoneme task (red), (ii) the differences between tasks in RT computed for each step (significant differences with adjustment for multiple comparisons marked by a red dot), (iii) the rate of change in RT measured between each successive pair (1st derivative) and (iv) the differences between tasks in the rate of change computed for each step (significant differences with adjustment for multiple comparisons marked by a red dot)

Row 4: Summary measures (left to right) - (i) the average reaction times per task and condition, (ii) the within and between tasks differences in average reaction times, (iii) the average rate of change, (iv) the within and between tasks differences in average rate of change.

Key: OS: original sounds, PE: f0-equalized sounds, TE: timbre equalized sounds. Bars represent the adjusted 95% CI on all graphs.

Reverse correlations

The average acoustic properties measured for the original sounds are displayed in Figure 3-10. As illustrated, ranking stimuli from male to female (gender task - top) or from /pæ/ to /tæ/ (phoneme task – bottom) gives different results. For instance focusing on the vowel, f0 is higher in the female stimuli (step 11 - /pæ/ female and /tæ/ female stimuli averaged) than in the male stimuli (step 1 - /pæ/ male and /tæ/ male stimuli averaged). In contrast, f0 does not change among the /tæ/ stimuli (step 11 - male /tæ/ and female /tæ/ stimuli averaged) and the /pæ/ stimuli (step 1 - male /pæ/ and female /pæ/ stimuli averaged). This is explained by the fact that we used 2 symmetric continua per subject. One continuum was going from male-/pæ/ to female-/tæ/ whilst the other was going from male-/tæ/ to female-/pæ/, and was 'reversed' in the phoneme task. This therefore cancels acoustic differences such as f0 observed in the gender task. By averaging acoustic properties across stimuli according to the PSE and by task, it was possible to highlight which acoustic features are distinctive between categories (for instance f0 allows to distinguish males from females but not /pæ/ from /tæ/) and within tasks. Note that this is different from looking at the extremes of the continua and comparing stimuli which would instead only reflect differences from the design. By taking the median difference of the average of stimuli above and below the physical middle (ideal listener) and the PSE (real subjects) we can however see if there is a difference between stimuli that can be diagnostic in the task at hand. It is also important to appreciate that despite supra-laryngeal filtering equalization (i.e. timbre), the vowels from the same and different speakers can have different formant values because the consonant environment influences the formant pattern in vowels (Hillenbrand et al., 2001).

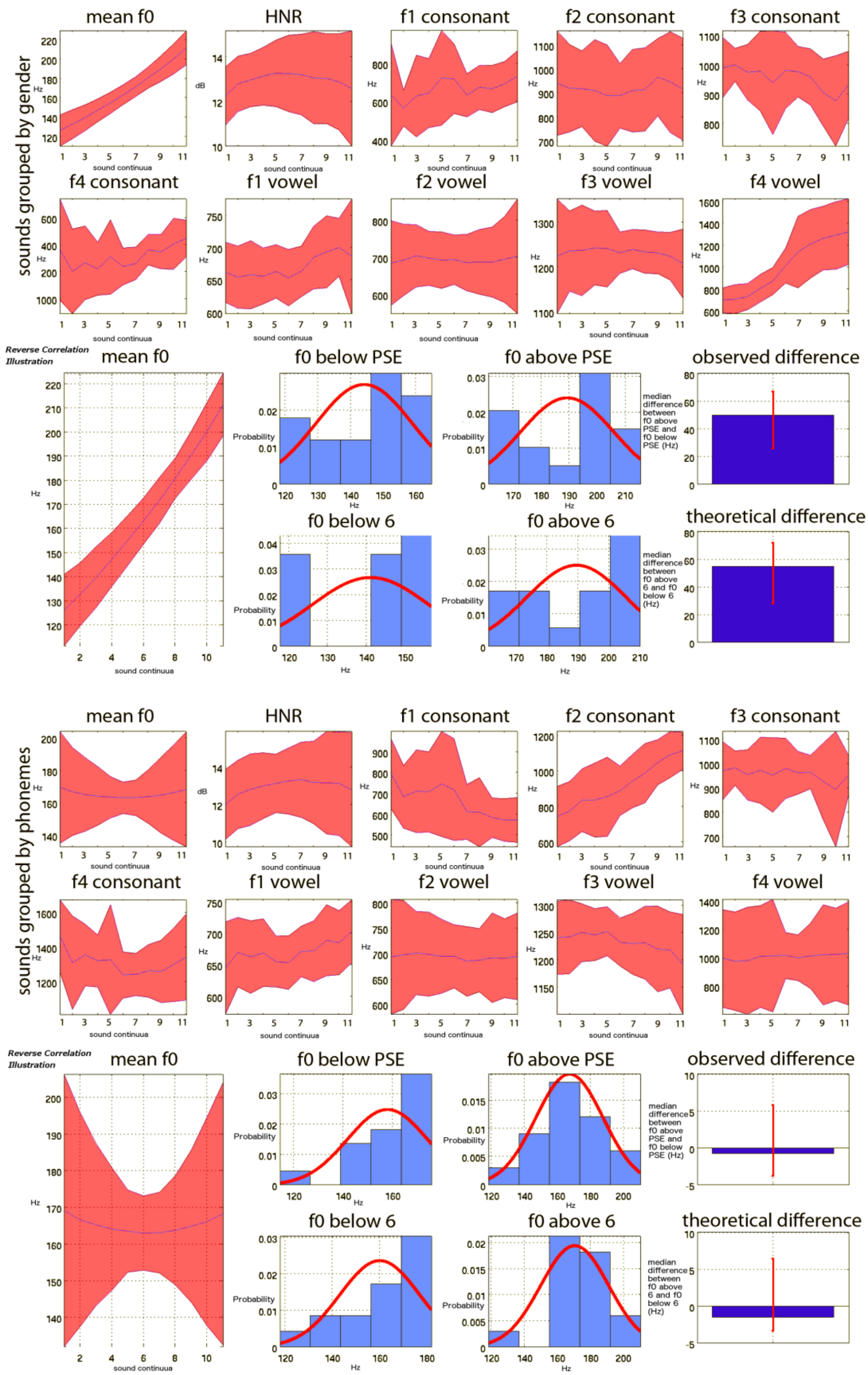

Figure 3-10: Mean values and 95% CI of acoustic properties measured on whole sounds (f0 and HNR) and on consonants and vowels separately (f1, f2, f3, f4).

Rows 1-2: Extracted acoustic properties averaged according to gender: therefore averaging of values across stimuli for each step from step 1 (all 100% male stimuli and therefore averaging male /pæ/ and /tæ/ together) to step 11 (all 100% female stimuli and therefore averaging female /pæ/ and /tæ/ together).
Row 3: Illustration of reverse correlation for f0: stimuli are separated as below or above the PSE for each subject or as below/above 6 (the actual physical middle) for the ideal listener and then averaged. Histograms show the distribution of f0 values for stimuli classified as males (below PSE or 6) and as females (above PSE or 6) separately. The median differences (bar graphs) are then computed, here showing higher f0 values in ‘females’ than ‘males’. These differences were also compared to each other to investigate whether subjects relied more or less on a given acoustic feature than the ideal listener (not shown here).
Rows 4-5: Extracted acoustic properties averaged according to phoneme: therefore averaging of values across stimuli for each step, from step 1 (all 100% /pæ/ and therefore averaging male and female /pæ/ together) to step 11 (all 100% /tæ/ stimuli and therefore averaging male and female /tæ/ stimuli).
Row 6: Illustration of reverse correlation for f0: stimuli are separated as below or above the PSE for subjects or above/below 6 (the actual physical middle) for the ideal listener and then averaged. Histograms show the distribution of f0 values for stimuli classified as /pæ/ and as /tæ/ separately. The median differences (bar graphs) are then computed, showing no differences in f0 here. These differences were also compared to each other to investigate whether subjects relied more or less on a given acoustic feature than the ideal listener (not shown here).

For the voice gender categorization task, comparisons of sound properties for original sounds categorized as ‘females’ had, as expected, a significantly higher fundamental frequency (mean f0) but also a higher f3-f4 formant dispersion on the vowel than stimuli categorized as ‘males’. These effects were observed for both the ideal listener and using subjects’ categorization performances. Comparison of the results from the ideal listener and from subjects’ categorization performances show, however, a smaller difference on f0 in our participants than expected (f0 difference [-5 -2] p=0, f3-f4 difference [-6 565] p=0.02). For f0-equalized sounds, reverse correlations based on the ideal listener and on subjects’ performances show that stimuli categorized as ‘female’ had a significantly higher f3-f4 formant dispersion on the vowel (Table 3-7), with a smaller difference for the observed than ideal differences (difference [-72 -33] p=0). Finally, for timbre equalized sounds, the reverse correlations on the ideal listener and subjects’ performances show that stimuli categorized as ‘female’ had significantly higher fundamental frequency (mean f0), f3-f4 formant dispersion on the consonant and f2-f3 formant dispersion on the vowel. In addition, a significantly higher HNR was also obtained, but only based on subjects' performances (Table 3-7). Comparisons between ideal and observed results revealed smaller differences on f0 and HNR in our participants than expected (f0 difference [-5 -2] p=0, HNR difference [-0.4 -0.1 p=0; f3-f4 difference for the formant dispersion on the consonant [-9 16] p=0.1 and f2-f3 difference for the formant dispersion on the vowel [-6 13] p=.15).

A		whole sound		consonant			vowel		
		f0	HNR	f1-f2	f2-f3	f3-f4	f1-f2	f2-f3	f3-f4
Male -Female	Observed	**[3466] p=0**	[-1.6 2.9] p=.02	[-43 166] p=.3	[-118-6] p=.02	[-247284] p=.17	[-123 134] p=.32	[-135 111] p=.14	**[61 712] p=0**
	Ideal	[3771] p=0	[-1.7 3.3] p=.3	[-86 180] p=.18	[-10129] p=.12	[-317306] p=.19	[-133 149] p=.31	[-149 128] p=.16	[49 756] p=0
/pæ/ -/tæ/	Observed	[-34] p=.46	[-0.3 1.4] p=.02	**[91 282] p=0**	[-4793] p=.32	**[-209-6] p=.003**	[-64 30] p=.35	**[-59-1] p=.003**	[-27 89] p=.007
	Ideal	[-36] p=.18	[-06 1.2] p=.03	**[167 321] p=0**	[-8750] p=.3	[-2310.1] p=.01	[-74 26] p=.19	**[-36- 0.1] p=0**	[-766] p=.004
B		whole sound		consonant			vowel		
		f0	HNR	f1-f2	f2-f3	f3-f4	f1-f2	f2-f3	f3-f4
Male - Female	Observed	[-0.1 0.03] p=.3	[-0.3 0.03] p=.3	[-39 179] p=.3	[-10555] p=.19	[-220219] p=.3	[-30 124] p=.7	[-136 36] p=.2	**[216 248] p=0**
	Ideal	[-0.2 0.4] p=.4	[-316] p=.05	[-135 193] p=.3	[-16098] p=.02	[-245514] p=.15	[-153 148] p=.3	[-146 144] p=.1	**[94 889] p=0**
/pæ/ -/tæ/	Observed	[-0.002 0.07] p=.09	[-0.14 1.2] p=.4	**[232 302] p=0**	**[-51-14] p=0**	**[-251-64] p=0**	[-66 45] p=.15	[-476] p=.3	**[5869] p=0**
	Ideal	[-0.1 0.3] p=.17	[-0.45 0.58] p=.2	**[145 289] p=0**	[-92179] p=.04	[-250289] p=.006	[-61 37] p=.19	**[-53 -2] p=0**	[-29 65] p=.006
C		whole sound		consonant			vowel		
		f0	HNR	f1-f2	f2-f3	f3-f4	f1-f2	f2-f3	f3-f4
Male -Female	Observed	**[4567] p=0**	**[0.9 3.8] p=0**	[-45 145] p=.16	[-13421] p=.3	**[92273] p=0**	[-184 116] p=.21	**[-182-16] p=0**	[-110 442] p=.41
	Ideal	**[27 71] p=0**	[-2.7 4.2] p=.003	[-153 148] 0.31	[-146144] p=.17	**[94889] p=0**	[-196 127] p=.13	**[-206-32] p=0**	[-91 462] p=.1
/pæ/ -/tæ/	Observed	[-25] p=.18	**[0.9 1.2] p=0**	[-425] p=.18	**[-44-21] p=0**	[-4858] p=.3	**[178 239] p=0**	**[-100-60] p=0**	[-0.6 263] p=.18
	Ideal	[-36] p=.18	[-0.91] p=.03	[-61 22] p=.17	**[-55-2] p=0**	[1654] p=.002	**[119 232] p=0**	**[-108-58] p=0**	[-30 140] p=.15

Table 3-7: Reverse correlation results for *A:* original sounds *B:* f0-equalised sounds *C:* timbre-equalised sounds.

In square brackets are the 95% CIs of the median difference between stimuli located above and below the PSE (observed) or above and below 6 (ideal). Significant p-values are marked in bold with grey background (alpha adjusted for multiple comparisons).

3.5.5. Discussion

Summary

Categorical responses were observed in all conditions with no *within* category differences (original vs. f0-equalised vs. timbre-equalised) seen in perceptual distances compared to sharp *between* category differences (gender vs. phoneme), where d' is different from 0 for ambiguous stimuli (higher perceptual distance and therefore rate of change). Comparison between tasks (gender vs. phoneme) revealed a higher rating in the phoneme task than in the gender task, especially in the timbre equalized condition. Between-task comparisons also showed faster processing in the phoneme task than in the gender task (the opposite of what was hypothesized). No effect of timbre equalization was observed in the gender task (on any measures), contrary to what has been previously reported (Pernet and Belin, 2012). Reverse correlations showed significant differences in *vowel* formant dispersions when stimuli were categorized as male vs. female, and significant differences in both *consonant* and *vowel* formant dispersions when stimuli were categorized as /pæ/ vs. /tæ/.

Is voice stripped away from speech?

It was expected that voice gender information would be processed faster than phonemic information in healthy participants, because of (i) a higher acoustic similarity between stimuli grouped by talker than grouped by phoneme and (ii) the hypothesized need for talker normalization. However, the opposite results were found i.e. faster RTs in the phoneme task than the gender task. If a normalization process was taking place during the phoneme task (as is a standard hypothesis), altering sounds so they had equalised f0 or equalised timbre should have led to faster RTs in those conditions, which was not the case. Together, these results contradict the hypothesis that voice information is stripped away or normalized to access phonemic content.

Attending to consonant vs. vowel

Do individuals use similar phonemic cues to identify both voices and words? Previous psycholinguistic studies suggest that this is the case (Remez et al., 1997), however the results presented in this study, with healthy controls, contradict this hypothesis. The reverse correlation analysis carried out shows that the gender task mainly relied on the vowel formant dispersions (as well as f0 when it was available), whereas the phoneme tasks (as expected) relied on the consonant formant dispersions. This therefore suggests that gender and phoneme are processed on the basis of different perceptual representations.

Conclusions

It is possible to conclude from the difference in reaction times observed in the phoneme task (faster) versus gender task (slower), along with the absence of effect of f0 or timbre equalization, that voice is not stripped away from speech to access phonemic content therefore challenging serial processing leading talker normalisation (voice processed first followed by speech – see Laing et al. (2012) for another challenge of this view).

It is also possible to show that different phonemic cues (as seen in the reverse correlation) are used to identify talkers as opposed to speech i.e. the use of the consonant formants in the phoneme task and the vowel formants in the gender categorization task therefore challenging the view of a common perceptual underpinning (i.e. acoustic cues are used to the same degree within perception of voice and speech) for speech and voice perception.

4. Chapter 4 – Intact and Impaired Speech and Voice Perception

In this follow-up study, the association/dissociation between voice and speech perception was investigated. The first aim was to investigate whether the same brain areas are used to process voices and speech, in particular, in particular the superior temporal gyrus vs. inferior frontal gyrus. A second objective was to investigate associations in language perception and/or production impairments (aphasia) with low level perceptual deficits (voice perception).

4.1. Impaired voice perception

There are only a limited number of studies that have looked at voice perception in people with aphasia, possibly because it is conceived as subordinate compared to speech perception (Belin et al., 2004). Early studies have reported cases of phonagnosia (Van Lancker et al., 1988; Van Lancker et al., 1989; Van Lancker and Canter, 1982), in which patients could not recognize familiar voices. These studies suggest an association with phonagnosia and right parietal lesions, supporting early hypotheses of right lateralization of voice perception and a neuro-functional dissociation. However, more recently discrimination of unfamiliar voices was further investigated and deficits were found to be associated with right or left hemispheric lesions in the temporal cortex, although some evidence also exists for voice deficit during fronto-temporal degeneration (Hailstone et al., 2010a; Hailstone et al., 2010b).

4.2. Association/ dissociation between impaired speech perception and aphasia

The inability to discriminate speech sounds has commonly been accepted as an underlying cause of aphasic auditory comprehension difficulties. However, auditory comprehension and its relation to phoneme perception in aphasia remains an inconclusive topic. Luria’s early study (1966) suggested auditory comprehension impairments in aphasia were due to phonemic perception difficulties. Conversely, a paper by Assal et al. (1976) presents a case study of receptive aphasia following left hemisphere damage, but with essentially intact speaker recognition, showing a neuroanatomical dissociation between voice and language. Blumstein (1977) supports this theory by suggesting that language comprehension impairments in aphasia do not correlate with perceptual deficits. Csepe et al.'s (2001) MMN (mismatch negativity component) study investigated the stages of impaired speech perception in people with aphasia, and showed that although phoneme discrimination tasks were more difficult for people with aphasia, they were not specific to the aphasia type. Van Lancker's studies (1988; 1989; 1982) supported these claims, suggesting no correlation between language impairment and voice perception. Recently Turkeltaub et al. (2012; 2010) reinforced this by stating that speech perception impairments do not necessarily predict comprehension deficits in people with aphasia.

4.3. Study 2 – Association/dissociation of voice and speech perception across the cerebral hemispheres investigated in intact and impaired processing

4.3.1. Introduction

Article:

Jones, A. B., Farrall, A., Belin, P., & Pernet, C. (2015). *Hemispheric Association and Dissociation of Voice and Speech Information Processing in Stroke*. Cortex, 71, 232-239.

Oral presentation:

Jones, A. *Dissociating voice from speech perception in stroke patients*. Paper presented at Scottish Stroke AHP Forum Conference; 2013 June 6; Perth, UK.

Poster presentation:

Pernet, C. R., Jones, A., Belin, P., & Farrall, A. *Dissociating voice from speech perception*. Poster session presented at SINAPSE Annual Scientific Meeting; 2013 May 15; Aberdeen, UK.

Contributions by Anna Jones: study design, ethical submission, patient recruitment, data collection and analysis, results interpretation, writing of the manuscript, writing and giving oral and poster presentations.

Speech perception is often viewed as ‘special’ (Liberman and Mattingly, 1989) because localised brain injury can elicit specific language impairments (e.g. aphasia), and healthy individuals are extremely efficient at categorizing phonemes and syllables despite large variations in the stimulus spectral patterns (Liberman et al., 1952). To achieve these performance levels, it has been hypothesized that this is due to a process known as ‘speaker normalisation’, whereby voice information (talker specific information) is extracted along with the speech signal, and then stripped away to access (invariant) phonemic content. This process is, however, challenged because general auditory learning mechanisms are capable of explaining category formation in the absence of invariant acoustic information. Birds can learn speech consonant categories with no obvious acoustic invariant cue (Kluender et al., 1987) and human listeners can readily learn non-speech categories that are similarly structured (Wade and Holt, 2005). In addition, several studies showed that talker variability influences speech perception. For instance the literature describes increased memory for words spoken by familiar voices, compared to non-familiar voices (Nygaard et al., 1994; Nygaard and Pisoni, 1996; Palmeri et al., 1993), and similarly enhanced discrimination of, and memory for, (non-familiar) speakers of our own language compared to speakers of another language (Language Familiarity Effect – (Perrachione and Wong, 2007)) even in the absence of intelligibility (Fleming et al., 2014). Most of these studies do not, however, specifically address the issue of phoneme perception, and thus acoustical regularity coming from multiple levels are at play.

Like speech perception, voice perception is often considered special (Belin et al., 2004; Scott, 2008). Humans easily recognize different voices, and this ability is of considerable social importance. Voice-selective areas have been demonstrated in the human brain (Belin et al., 2000; Belin et al., 2002; Pernet et al., 2015a), localized bilaterally along the upper bank (middle and anterior) of the Superior Temporal Sulcus (STS) (Alho et al., 2006; Belin et al., 2000) and also in the inferior and orbitofrontal cortex (Charest et al., 2012; Fecteau et al., 2005) as well as surrounding insular cortex (Johnstone et al., 2006; Rama et al., 2004). This neural selectivity for voice has also been established in other mammals, in particular primates (Johnstone et al., 2006; Rama et al., 2004) and more recently dogs (Andics et al., 2014). Given the presence of con-specific voice neural selectivity in these animals, we can establish that neural voice selectivity is an old evolutionary feature (~100 million years for a common ancestor between humans and dogs, ~25 million years for a common ancestor between macaques and human), preceding the appearance of speech (~5 to 2 million years ago for proto-language and ~150,000 to 300,000 years ago for speech (Perreault and Mathew, 2012)). Following psycholinguistic studies suggesting that phonetic attributes are an intrinsic component of the perception and identification of individuals, i.e. the recognition of voices (Remez et al., 1997), it is possible that some brain regions dedicated to speech co-opted neurons already involved in con-specific voice processing.

In this study, the relationship between the processing of phoneme and talker information was investigated, by comparing performances of right fronto-temporal (non-aphasic), left fronto-temporal aphasic and left fronto-temporal non-aphasic stroke patients. Each participant categorized sounds from pitch equalized morphed continua as being male-female or /pae/-/tae/ (Pernet et al., 2013). Stimuli were the same in both tasks, and participants had therefore to discard talker specific or phoneme specific information depending on the task at hand. Given the importance of the right STS (Bestelmeyer et al., 2011) and right Middle and Inferior Frontal Gyrus (MFG-IFG) (Charest et al., 2012) in talker information processing, it was hypothesized that right hemisphere patients will show a dissociation between the two tasks. In contrast, following the hypothesis of co-optation of voice selective neurons in phoneme processing, it was hypothesized that left hemispheric aphasic patients would not show such dissociation, while non-aphasic patients would be impaired for voice but not phoneme.

4.3.2. Method

The experiment used runs in Matlab with the psychophysical toolbox (Brainard, 1997; Kleiner et al., 2007). The study was approved by the NHS Lothian South East Scotland Research Ethics Committee 01 *(REC reference number: 11/SS/0055)* and NHS Lothian Research and Development *(R&D project number: 2011/W/NEU/09).*

Participants

Twenty-five stroke patients (14 males, 11 females) with a median age of 69 years (min 39, max 85) were recruited into the study. At the time of testing, all patients were at the chronic stage (median time between stroke and testing 90 ± 17 days). Participants were recruited as inpatients and outpatients from Lothian NHS hospitals via stroke physicians and Speech & Language Therapists between 10 and 60 weeks post-stroke with the sole inclusion criterion of a stroke affecting perisylvian tissues (Table 4-1). Exclusion criteria were the presence of a previous stroke and/or English not being the participant's first language.

Consent was sought from all participants to review their medical & Speech and Language Therapy records (if appropriate), carry out formal assessments of language, participate in voice and speech perception tasks, and to review and use clinical brain scans (if appropriate) for lesion-symptom mapping. Participation in the study required consent to all aspects.

All participants were provided with detailed information regarding the nature of the study and their potential role in it. Information sheets and consent forms were provided in an aphasia friendly format (see Appendix 3 & 4). Once participants had consented to take part in all aspects of the study, a letter was sent to their G.P informing them of their participation and providing information on the study (see Appendix 5).

Subject	Stroke type	Arterial territory	Regions of interest showing lesion	WAB AQ and classification
1	Ischemic	Left MCA 1,2,3,4,5,6	'IFG' 'insula' 'anterior STS' 'posterior STS' 'Heshl gyrus'	19 - Global
2	Ischemic	Left MCA 1,2,4,5	'IFG' 'insula' 'Heshl gyrus'	38.15 -Transcortical Motor
3	Ischemic	Left MCA 1,4	'IFG' 'MFG'	60.9 - Broca's
4	Ischemic	Left MCA 4,6	'IFG' 'MFG'	89.6 - Anomia
5	Haemorrhagic	Left MCA 1,2,3,6	'IFG' 'anterior STS' 'posterior STS'	17 - Wernicke's
6	Ischemic	Left MCA 1	'IFG' 'insula' 'Heschl gyrus'	36.6 - Broca's
7	Ischemic	Left MCA 1	'IFG' 'insula'	20.8 - Brocas
8	Ischemic	Left MCA 2, 3,6	'IFG' 'insula' 'anterior STS' 'posterior STS' 'Heschl gyrus'	62.9 - Wernicke's
9	Ischemic	Left ACA	'IFG'	83.6 - Anomia
10	Ischemic	Left MCA 2,3,6	'IFG' 'insula' 'anterior STS' 'posterior STS' 'Heschl gyrus'	65.1 - Conduction
11	Ischemic	Left MCA 5	'IFG'	99.4
12	Ischemic	Left LSA	'IFG' 'insula'	97.9
13	Ischemic	Left PCA	'insula' 'Amygdala-Hippocampal complex'	99.4
14	Ischemic	Left MCA 4,5,6	'IFG'	100
15	Ischemic	Left MCA 1	'IFG' 'insula' 'Heschl gyrus'	99.4
16	Ischemic	Left LSA	'insula'	100
17	Ischemic	Right MCA 1,4	'IFG'	98.4
18	Haemorrhagic	Right MCA 1,4,5	'IFG' 'insula'	98.8
19	Ischemic	Right MCA 2,4,5	'IFG' 'insula' 'anterior STS' 'posterior STS' 'Heshl gyrus'	98
20	Ischemic	Right LSA	'IFG'	98.8
21	Ischemic	Right MCA 2, 5	'IFG' 'insula' 'anterior STS' 'posterior STS' 'Heschl gyrus'	100
22	Haemorrhagic	Right MCA 5	'IFG'	100
23	Ischemic	Right MCA 1	'IFG' 'insula'	98.6
24	Ischemic	Right MCA 1,2, 3, 4, 5, 6	'IFG'	98.8
25	Ischemic	Right MCA 1	'IFG' 'insula'	93

Table 4-1: Patient clinical scan evaluations based on the International Stroke Trial III (Whiteley et al., 2006) along with the Western Aphasia Battery Aphasia Quotient (WAB AQ) and classification (Kertesz, 1982).

ACA stands for anterior cerebral artery. LSA stands for lenticulo-striate. IFG stands for inferior frontal gyrus. STS stands for superior temporal sulcus. PCA stands for posterior cerebral artery. MCA stands for middle cerebral artery, and 1 to 6 refers to sub-territories (1 small cortical, 2 basal ganglia, 3 lateral to the ventricle, 4 anterior half of peripheral MCA, 5 posterior half of peripheral MCA, 6 whole peripheral MCA).

All patients were tested for their mood (Visual Analogue Self Esteem Scale – VASES (Brumfitt and Sheeran, 1999) and The Depression Intensity Scale Circles – DISCSs (Turner-Stokes et al., 2005)) and language abilities (Western Aphasia Battery – WAB (Shewan and Kertesz, 1984)). No patient had language deficits in the group with right hemisphere lesions (N=9, 5 males and 4 females, WAB median score 98.8), and 10 out of 16 patients showed signs of aphasia in the left hemisphere group (N=10, 5 males and 5 females, WAB median score 49.5 for aphasics vs. N=6, 4 males and 2 females, WAB median score 99.4 for non-aphasics – percentile bootstrap difference 49.8 [15 80] p=0). Kruskall-Wallis ANOVA showed that groups did not differ in terms of median age ($\chi^2(2,22)=4.58$ p=0.1), in median time delay between stroke and testing ($\chi^2(2,22)=1.68$ p=0.43) or depression scores ($\chi^2(2,22)=4$ p=0.13 for VASES and $\chi^2(2,22)=2.19$ p=0.33 for DISCS).

Table 4-2 provides a summary of all demographic and test results.

	Left Aphasic	**Left Non Aphasic**	**Right Non Aphasic**
Sample size	10	6	9
Age (years)	71 ± 3.8	61 ± 8.34	65 ± 4.7
Stroke to Test time (days)	89 ± 18	133 ± 83	86 ± 58
Tone discrimination threshold (Hz)	8.25 ± 1.16	5.37 ± 2.31	6.75 ± 1.09
WAB score (out of 100)	49.5 ± 14	99.4 ± 0.5	98.8 ± 0.3
VASES (out of 50)	35 ± 3.4	44.5 ± 2.9	37 ± 4.9
DISCS (out of 4)	2.5 ± 0.4	2 ± 0.39	3 ± 1

Table 4-2: Medians with standard errors of patient demographics, basic auditory, language, and depression characteristics.

Stimuli

The experiment used exactly the same pitch-equalized stimuli as presented in Chapter 3 (Pernet et al., 2013). Previously it was thought that pitch was an important cue used in both voice and phoneme perception. However, more recently it has been recognised that other cues are actually more important in the perception of voices and phonemes. For example, phoneme categorization, in English, has been shown to rely on acoustic cues such as voice-onset-time (VOT) and formant transitions (Koch et al., 1999). Gender perception is affected by the size of the larynx and vocal tract (Belin et al., 2004; Lass et al., 1976) and gender is perceived using both pitch (Landis et al., 1982) and timbre (Bachorowski and Owren, 1999). However, because of the pitch overlap in the population between males and females (Titze, 1994), pitch alone can be unreliable for gender categorization (Hanson and Chuang, 1999). Previous studies have argued that pitch (f0) and formants are the most salient cues to distinguish speaker's sex in the context of vowels (Whiteside, 1998) with a major role of pitch (Gelfer and Mikos, 2005). The study presented in Chapter 3 (Pernet et al., 2013) showed that timbre equalization of the vowel did not impair performance, as had been previously observed (Pernet and Belin, 2012). Rather it was observed that voice gender categorization could be performed accurately using timbre information only (i.e. with pitch equalized stimuli), thus demonstrating that pitch is not mandatory to gender categorization. Reverse correlation results also indicated that formants on the vowel were a major feature in distinguishing male from female stimuli (see also: (Ghazanfar and Rendall, 2008; Rendall et al., 2005). Together, these results demonstrate a predominant role of timbre, over pitch, when carrying out gender categorization in the context of phonemes, with formants rather than pitch acting as the major cues. Therefore, given the lack of importance of pitch in both phoneme and gender perception it was decided that pitch equalised stimuli would only be used.

Paradigm

The experiment was identical to that presented in Chapter 3 (Pernet and Belin, 2012), except that only pitch equalized stimuli were used (as detailed). Participants were presented with two 2 alternative forced choice identification tasks: voice gender (male vs. female) and phoneme (/pae/ vs. /tae/), and responded by button press on a keyboard. For each task, the same two continua of morphed sounds were used: the 1st continuum going from a Male-/pae/ to a Female-/tae/ and the 2nd continuum with the same speakers going from a Male-/tae/ to Female-/pae/. Participants heard each stimulus in pseudo-random order (e.g. 100% Male-/pae/, 90% Male-/pae/ and 10% Female-/tae/, 80% Male-/pae/ and 20% Female-/tae/, etc. until 100% Female-/tae/) six times each. In total, each participant heard 132 stimuli (2 continua * 11 steps * 6 trials) per task. This design allowed investigation of the effect of the task while controlling for the general acoustic characteristics of the stimuli, since the same stimuli were used in both tasks. Eighteen different continua of stimuli were generated from 6 different speakers (3 males and 3 females pronouncing /pae/ and /tae/) and randomly assigned to participants. Task order and key orientation were counterbalanced between participants.

Between each task, an interfering tone discrimination task was also performed. In this task participants heard pure tones of various frequencies corresponding to the male and the female ranges and were required to state if 2 consecutive sounds were the same or different. The task followed a 2 down, 1 up step-wise procedure (Levitt, 1971) such that all participants had similar performances (70.71% correct). Importantly, a number of previous studies claim to show dissociated speech and voice perception (Aaltonen et al., 1993; Binder et al., 2000; Celsis et al., 1999), however the use of pure tones as stimuli to investigate voice perception in such studies is not satisfactory. It is known that pure tones cannot be used as voice stimuli as they only require simple auditory processing areas of the brain i.e. planum temporal & A1 (Rauschecker and Tian, 2000). Therefore this task was primarily designed to minimise the influence of one categorisation task on the other, but also allowed control for basic auditory impairments. No significant differences were observed between groups on this task (Kruskall-Wallis ANOVA, $\chi 2$ (2,22)=1.93 p=0.38,
Table 4-2).

4.3.3. Data Analysis

Behavioural Classification

To assess the independence of phoneme and gender categorization tasks, behavioural performances were binarised as impaired vs. unimpaired. Therefore, for each subject, labelling curves (percentage of female response or percentage of /ta/ responses) were obtained by averaging repeated trials from the different continua (Figure 4-1). Each participant was then classified as being impaired or unimpaired based on his or her ability to perform outside chance level, at least one time for the first 3 stimuli and at least one time for the last 3 stimuli along the sound continua. This implies that if a participant answered correctly for at least one of the initial stimuli (100% or 90% or 80% male or /pae/) and one of the final stimuli (100% or 90% or 80% female or /tae/), he or she was considered unimpaired.

It is important to note that repeated analyses using an incremental classification criterion were also performed, to obtain the most robust criterion levels. However, changing the classification criterion to the inability to perform above chance for extreme stimuli only, then for at least 1 of 2, at least 1 of 3, at least 1 of 4, and at least 1 of 5 extreme stimuli – all gave similar results (20/25, 19/25, 19/25, 18/25, 18/25 classified as impaired, respectively). The number of patients showing dissociations was relatively stable i.e. associations/dissociations results were unchanged by changing the criterion. In left aphasic group, out of the 10 patients, the number of subjects showing dissociations was 4, 4, 5, 5, and 5, respectively. In the left non-aphasic group, the same 2 patients always show a dissociation. In the right hemisphere group, out of 9 patients, the number of subjects showing dissociations was 7, 7, 7, 6, and 6, respectively. Therefore, the classification system used was based on each participant’s ability to perform outside chance level at least once for the first 3 stimuli and at least once for the last 3 stimuli along the sound continua.

In normal healthy participants, using this criterion, unimpaired results are achieved very easily. Using the data presented in Chapter 3 (Pernet et al., 2013), 100% of healthy controls (N=18, 9 males, 9 females) were unimpaired on all the pitch-equalised categorization tasks completed (Figure 4-2). From the resulting classification, the independence between the phoneme and gender categorization performance was tested for each group using a McNemar test with exact central probability (Fay, 2010).

Lesion to symptom mapping

Each participant received one or more CT or MRI scans from the National Health Service due to the nature of their presenting symptoms, and the scan the closest in time to the subject's participation in this study was reviewed. One-to-one mapping was computed using a McNemar test with exact central probability (Fay, 2010) between behavioural deficits (impaired/non-impaired) and 6 Regions Of Interest (ROI - lesioned/non-lesioned). Significance was considered at alpha=0.0083% i.e. Bonferroni corrected for the 6 ROI. The ROI classification was performed by an expert neuro-radiologist, Prof. Andrew Farrall, who followed a detailed protocol (see Appendix 6) which was adapted from the International Stroke Trial III (Whiteley et al., 2006). ROIs considered were the middle/inferior frontal gyrus (involved in gender categorization (Charest et al., 2012; Fecteau et al., 2005)), Heschl's gyrus (also called Transverse Temporal gyrus – involved in language learning and spectral/pitch information processing (Warrier et al., 2009)), the superior temporal gyrus (anterior/posterior, involved in general auditory processes but also voice perception (Belin et al., 2004)), the insula (involved in central auditory functions, in particular temporal resolution and sequencing (Bamiou et al., 2006)) and the amygdala-hippocampal complex (involved in memory and emotional voice perception (Johnstone et al., 2006; Rama et al., 2004)).

Behavioural quantitative analyses

Each participant's ability to distinguish between male-female and /pae/-/tae/ stimuli was investigated using signal detection theory (Macmillan and Creelman, 2004), similar to Chapter 3 (Pernet et al., 2013), computing the perceptual distance between (i) successive pairs of stimuli along each continua (d') and (ii) the extreme stimuli of each continuum (global d' - the distance between 100% male or /pae/ and 100% female or /tae/). Categorical perception can be described as the division of a continuum of stimuli by a boundary, whereby stimuli on one side of the boundary are perceived as belonging to one category and stimuli on the other side are perceived as belonging to another category, and perceptual distances within a category are low. Here categorical perception was investigated in each group and task by testing if d' values differed from 0 (bootstrap t-test with Bonferroni correction (Wilcox, 2012)) for each pair of stimuli along the continuum, thus indicating exactly a perceptual boundary. Finally, differences between gender and phoneme categorization performances were tested comparing global d' values within groups (percentile bootstrap of the difference with adjustment for multiple testing) and compared to the healthy participant data in Chapter 3 (Pernet et al., 2013) (mean difference between groups, with adjustment for multiple comparisons based on the maximum statistics (Wilcox, 2012)).

4.3.4. Results

Behavioural classification

Of the 25 patients recruited, 19 showed at least one categorization deficit (Table 4-3), which was defined as the inability to perform beyond chance level for at least one of extreme stimuli (100%, 90% or 80% male/'pa', female/'pa', male/'ta' or female/'ta'). As hypothesised, a significant dissociation in right fronto-temporal patients with impaired voice gender categorization vs. intact phonological categorization was observed (8 out of 9 patients, χ^2=7, p=0, Φ=inf). In left fronto-temporal patients, no dissociation was found for both the aphasic and the non-aphasic groups. Aphasic patients tended to show both phonological and voice gender categorization deficits (χ^2=0.2, p=1, Φ=0.6, Figure 4-1, Figure 4-3 & Table 4-3) and non-aphasic patients tended to perform normally in both tasks (χ^2=0, p=0.5, Φ=1, Table 4-3). The same association/ dissociation patterns were observed when varying the categorisation deficit criteria. Figure 4-1 shows average response proportion curves for each group of subjects.

			Gender task	
			Unimpaired	**Impaired**
Phoneme task	**Left fronto-temporal aphasics**	**Unimpaired**	1	2
		Impaired	3	4
	Left fronto-temporal non-aphasics	**Unimpaired**	4	1
		Impaired	1	0
	Right fronto-temporal non-aphasics	**Unimpaired**	1	7
		Impaired	0	1

Table 4-3: Contingency table showing the classification of patients (per group) for the gender and the phoneme task.

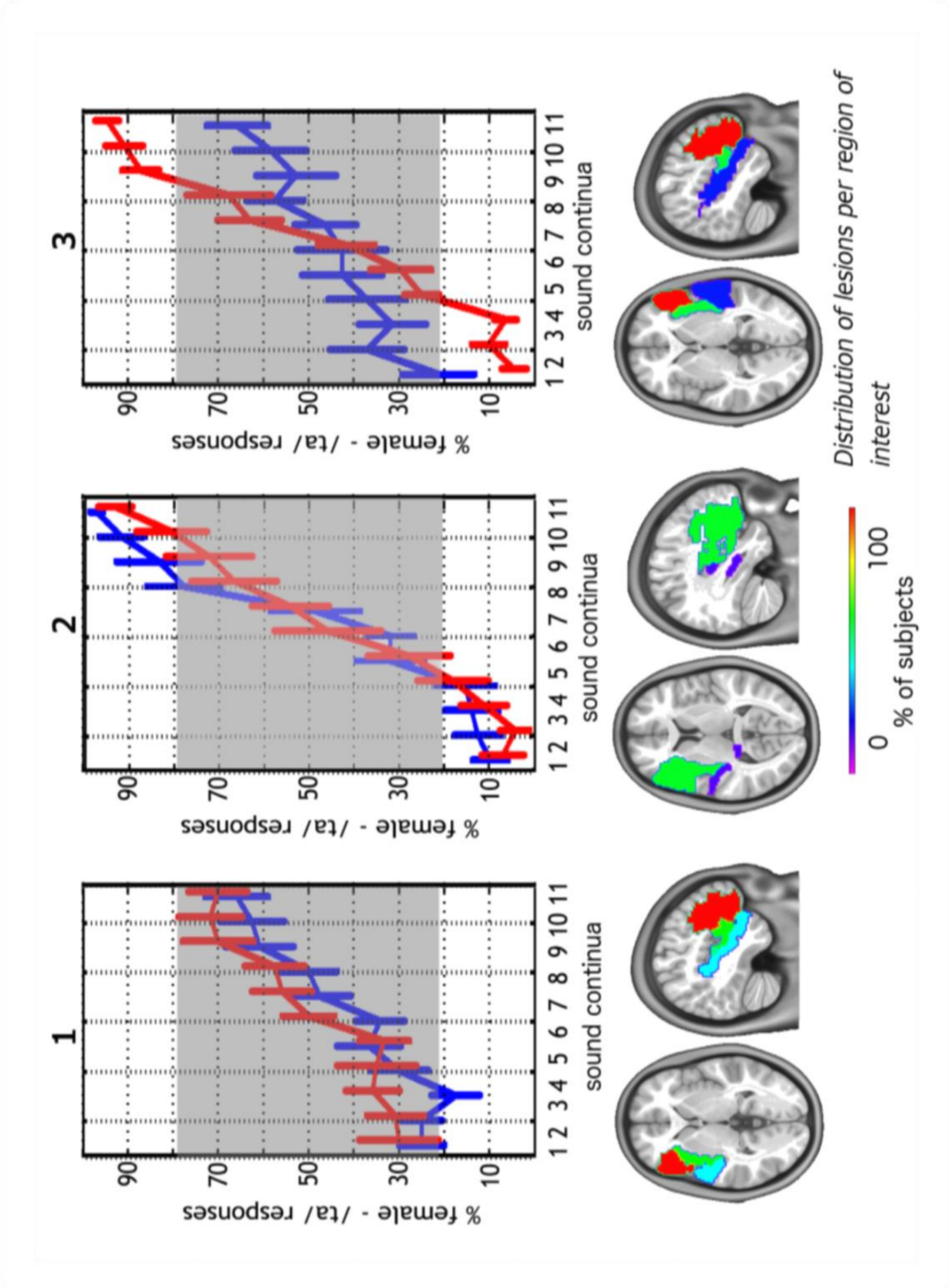

Figure 4-1: Average response labelling curves for each group of subjects

1 – subjects with left hemispheric damage & associated aphasia; 2 - subjects with left hemispheric damage & no associated aphasia; 3 - subjects with right hemispheric damage & no aphasia. Red lines represent phoneme task; blue lines represent gender task, bars represent bootstrap standard errors, shaded areas indicate chance level from 21.09% to 78.91%. The bottom row shows the distribution of lesions (0 -100%) per ROI, for each group, superimposed onto the MNI 152 template.

Lesion to symptom mapping

One-to-one mapping between the behavioural classifications (impaired / unimpaired) and region-of-interests (lesioned / not-lesioned) showed that gender categorization impairments are associated with right frontal lesions (χ^2=8, p= 0.0078 – see Figure 4-2 for details). No other ROI showed any significant results.

Behavioural quantitative analyses

Analyses of perceptual distances (d') between successive pairs of items along continua revealed that none of the patient groups had increased perceptual distances for ambiguous items (Figure 4-4), contrary to healthy subjects as shown in Chapter 3 (Pernet et al., 2013). This result indicates a generalized reduction in categorical boundaries following stroke. Further analyses on global perceived distances (i.e. the d' computed between 100% male and 100% female or 100% /pae/ and 100% /tae/ stimuli) showed that, compared to controls, right fronto-temporal patients had a lower global d' in the gender categorization task only (difference between controls vs. patients= 1.3, adjusted 95% CI [0.8 1.8] for Male/Female & difference between controls vs. patients= 0.06, adjusted 95% CI [-0.2 0.3] for 'pa'/'ta'). For left fronto-temporal aphasic patients, global d' was lower than in controls for both male/female stimuli (difference between controls vs. patients= 1.4 adjusted 95% CI [0.5 2.2]) and pa/ta stimuli (difference between controls vs. patients= 1.4, adjusted 95% CI [0.4 2.4]). For left fronto-temporal non-aphasic patients, global d' did not differ from controls (difference between controls vs. patients= 0.11 adjusted 95% [-0.27 0.49] for Male/Female & difference between controls vs. patients= 0.17, adjusted 95% CI [-0.49 0.83] for Pa/Ta). Finally, looking at within group differences, whilst right fronto-temporal patients show a significant difference between tasks, no significant differences were observed for both left fronto-temporal groups (Figure 4-1 and Table 4-4), which concurs with the classification results.

	Controls	**Left fronto-temporal subjects with aphasia**	**Left fronto-temporal subjects without aphasia**	**Right fronto-temporal subjects (no aphasia)**
Distance Male/Female	2.54	1.12	2.43	1.23
Distance Pa/Ta	2.55	1.1	2.17	2.49
Difference and adjusted 95% CI	0.01 [-0.1 0.1] p=0.872	0.019 [0.9 0.8] p=0.958	0.054 [-0.4 0.6] p=0.86	1.25 [1.8 0.7] p=0

Table 4-4: Global perceived distance (d’) computed for each group with within group differences. Adjusted 95% confidence intervals and p values.

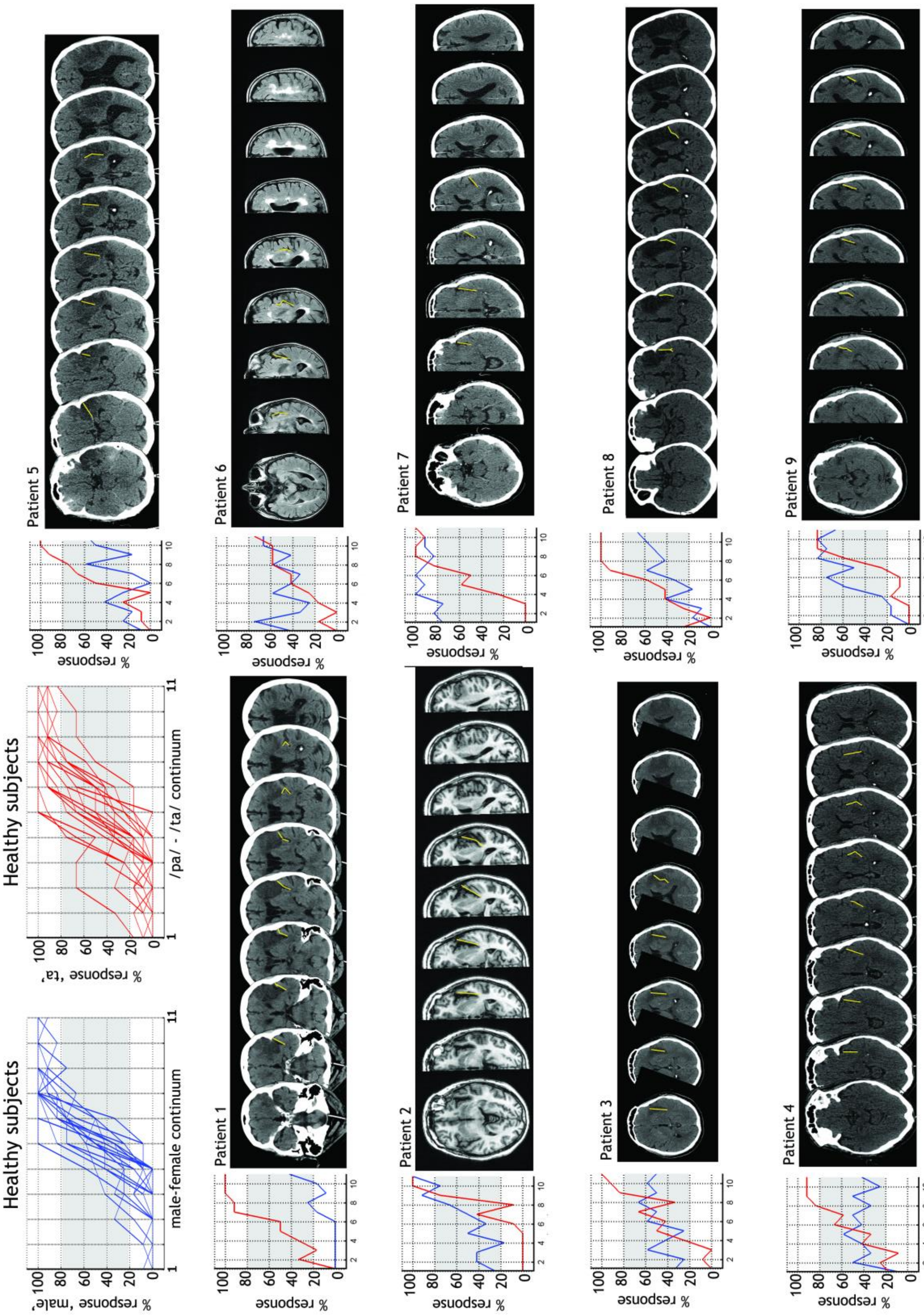

Figure 4-2: Labelling curves/ classification system for normal healthy subjects and each right hemisphere damaged patient, along with associated patient axial scans.

Top left: two labelling curves for responses in 18 normal healthy subjects ((Pernet et al., 2013)/ Chapter 3), showing that the classification system is successful, with controls scoring 100% correct on both gender and phoneme tasks. Below: the same labelling curves (red-phoneme; blue-gender) for each right hemisphere patient (1-9) are shown, and the associated axial slices showing the patient's lesion in relation to the Sylvian fissure (highlighted in yellow). Importantly, the STG is intact in 7 out of 9 cases (patients 3 and 5 have frontal & temporal lesions), suggesting this region is not critical for gender/voice categorization.

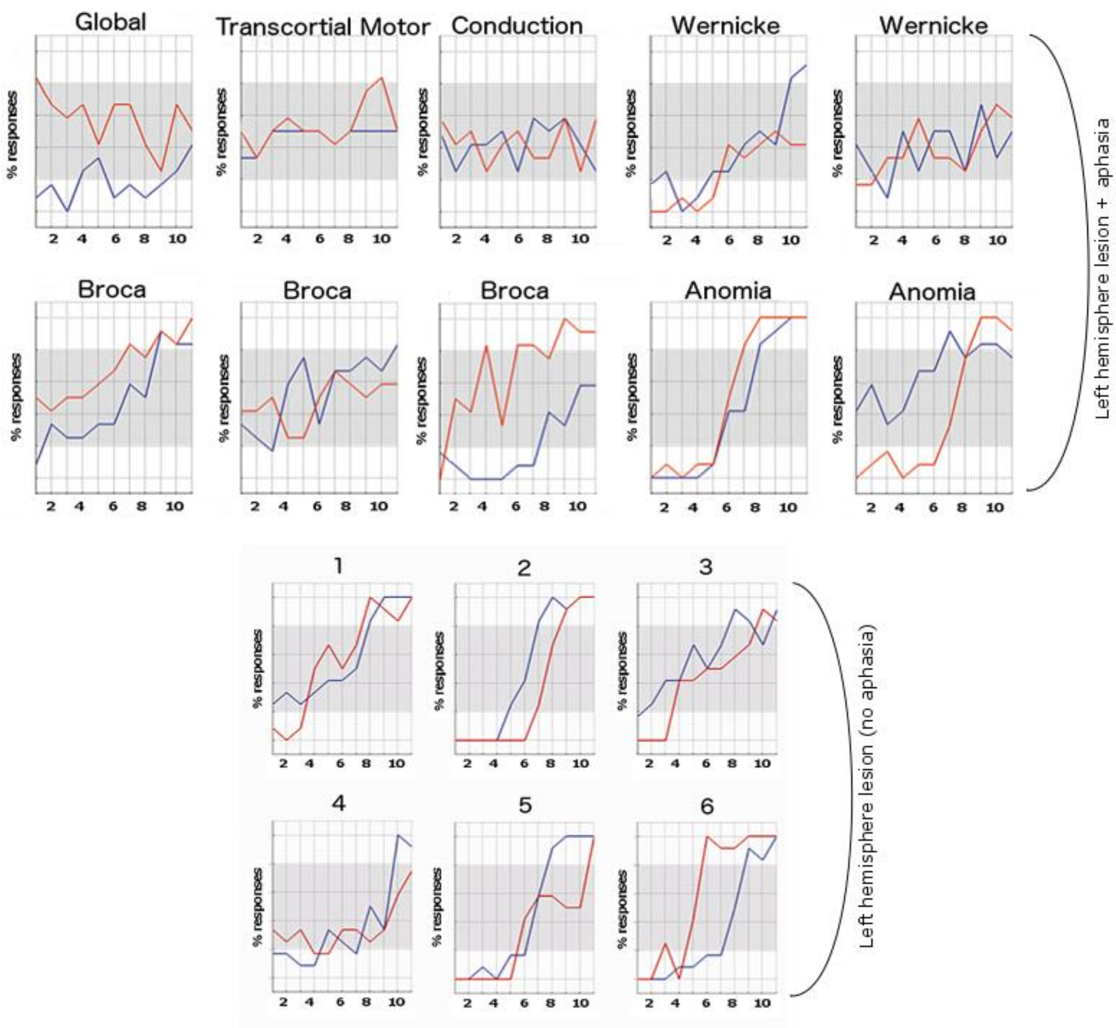

Figure 4-3: Response curves with classification criteria banding for patients with left hemisphere lesions.

Red =phoneme; blue=gender. Rows 1 & 2: 10 left hemisphere patients with aphasia, labelled according to the aphasia classification. Rows 3 & 4: left hemisphere patient without aphasia, labelled according to patient number 1-6.

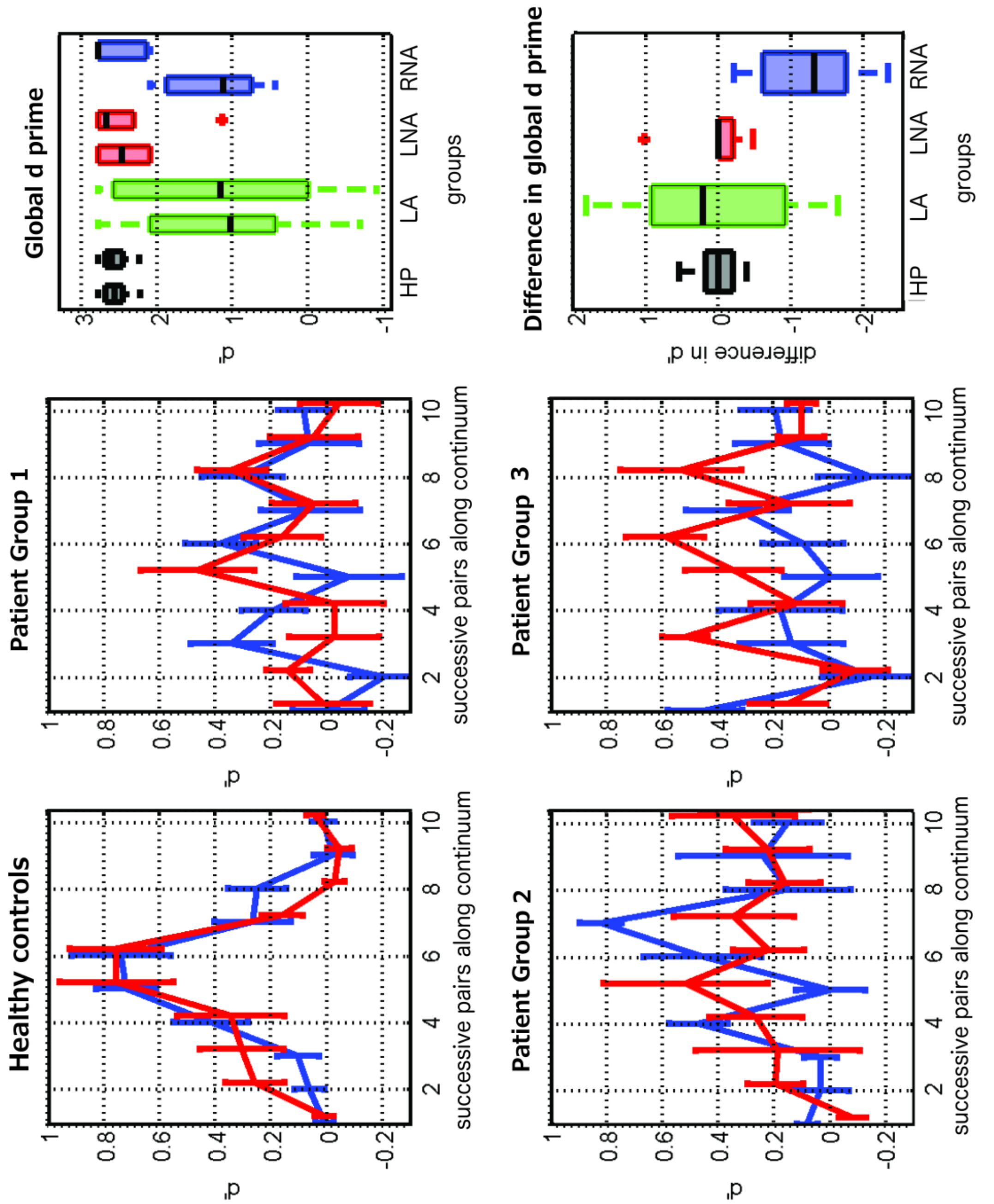

Figure 4-4: Perceptual and global perceptual distances (d prime (d')).

Curves show d' values in the gender categorization task (blue) and the phoneme categorization task (red) for the healthy controls and 3 patient groups: 1 – Patients with left hemispheric damage and associated aphasia; 2 – Patients with left hemispheric damage and no aphasia; 3 – Patients with right hemispheric damage and no aphasia. Bars represent bootstrap standard errors. Box plots show the global d' (top) and the difference in global d' between the two tasks (bottom) for control healthy participants (HP - black), left aphasic (LA - green), left non-aphasic (LNA - red) and right non-aphasic (RNA - blue) patients. Boxes represent the interquartile range with the median, and whiskers show the 1.5*interquartile distance with outliers marked beyond that points.

4.3.5. Discussion

Brain/behaviour association

Based on functional MRI (Belin et al., 2002; Charest et al., 2012) and Transcranial Magnetic Stimulation (Bestelmeyer et al., 2011) results observed in healthy volunteers, it was hypothesized that patients with right fronto-temporal stroke would show a deficit in gender categorization but intact phonological performances. The results, using both a qualitative (classification of percentage of responses) and a quantitative (signal detection theory) approach, confirm this hypothesis. To the best of found knowledge, this is the first time that such a deficit has been described in the literature. All of the patients presenting with right frontal lesions showed a deficit in voice categorization, thus demonstrating a significant brain/behaviour association. Previous studies have reported cases of phonagnosia (Van Lancker et al., 1988; Van Lancker et al., 1989; Van Lancker and Canter, 1982) in which patients could not recognize familiar voices; but this deficit was associated with right parietal lesions. When discrimination of unfamiliar voices was tested, deficits were associated with temporal (left or right) lesions although some evidence also exists for voice deficit during fronto-temporal degeneration (Hailstone et al., 2010a; Hailstone et al., 2010b). What remains unclear is 1) if the deficit is specific to gender categorization or if it also relates to identity, and 2) what is the role of the right IFG. On one hand, studies comparing attention to voice vs. speech found voice specific effects over the right STS (von Kriegstein et al., 2003), and this effect was related to the speaker identity (Schall et al., 2014). On the other hand, using a continuous carry over-over design allowing acoustic from perceived distances to be distinguished during gender categorization, Charest et al. (2012) showed that the STS processes gender related acoustic information (and thus identity as well) whilst the right IFG is involved in perceived gender related distances. In most studies, stimuli are pitch equalised. It has previously been shown that equalising pitch does not influence performance or reaction times (Pernet et al., 2013; Pernet and Belin, 2012). Therefore timbre, and consequently spectro-temporal analysis, is a key element in gender categorisation. There is no doubt that the aphasic patients in this study found the gender categorization task difficult, even when the anterior STS was intact (Figure 4-2 – patients 1, 2, 4, 6, 7, 8, 9). This difficulty was not related to the absence of pitch information since their pitch perception threshold did not differ from other patient groups. It has been proposed that the right IFG plays a general role in voice recognition and social communication at large, since direct connections have been demonstrated between the STG and the ventrolateral prefrontal cortex of the macaque (equivalent of IFG in humans) along with vocalisation responsive cells (Romanski, 2012). Voice deficits observed here, following right IFG lesions, are often also associated with insula lesions (although the association is not significant), which could then reflect a disconnection syndrome from the STS.

Link between voice and phoneme perception

The second main finding in this study is the close relationship between voice (gender categorization) and speech (phoneme categorization) in left hemisphere patients with half of the patients (9 out of 16) either impaired or unimpaired in both tasks. The results observed in left hemisphere damaged patients were heterogeneous. The absence of one-to-one mapping between speech perception deficits and brain lesions is not completely surprising since many studies have found that phonological categorization depends critically on the left supramarginal gyrus, a region not investigated in the current study.

Phonological deficits were also not always associated with aphasia (4 out of 16 patients), which concurs with the idea that sub-lexical speech perception impairments do not necessarily predict auditory comprehension deficits (Turkeltaub and Coslett, 2010). More than half of the left hemisphere damaged patients (9/16) were either impaired or unimpaired in both tasks. In the aphasic group, 4 out of 10 patients showed a double deficit and 5 others showed reduced performances, resulting in lower perceived distances between extreme stimuli (either male-female, or /pae/-/tae/). One possible explanation is that aphasic patients simply did not understand the instructions. Because these same patients could, however, perform other tasks with more complex instructions (e.g. the tone discrimination task used to check their pitch perception - see methodology section), this pattern of results can be considered as supporting the co-optation hypothesis. Indeed, if we accept that there is speech-selectivity in the left fronto-temporal cortex, then we have to conceive that this selectivity can be associated with the mechanisms that produce and perceive the sounds of speech (McGettigan and Scott, 2012). Such a relationship can also be found through audio-visual integration. For example, in the McGurk effect (McGurk and MacDonald, 1976), listeners combine auditory and visual information to come to a phonetic judgment about what is being said. Yet audio/video integration is blocked (i.e. the effect disappears or is attenuated) when listeners know the talkers (Walker et al., 1995), supporting the idea that voice and phoneme perception are linked.

Inter-hemispheric processing

The dissociations observed in 3 left hemisphere patients, also reveals that gender categorisation deficits can be observed following left hemisphere lesions, and therefore that gender categorisation is processed bilaterally. Voice processing is carried out bilaterally, with a special role of the right frontal areas. The data presented in Chapter 3 (Pernet et al., 2013) shows significant differences in reaction times between phoneme and gender tasks in healthy controls (RTs significantly shorter in phoneme compared to gender), contradicting the hypothesis of speaker normalization. In the light of the results of the current study it is now proposed that the reason behind this extended reaction time is that gender categorisation requires more inter-hemispheric communication, whereas phoneme processing can be carried out solely using the left hemisphere.

Further investigations

Interestingly, during the phoneme task a number of patients anecdotally reported perception of /kae/ with stimuli presented at the centre of the /pae/-/tae/ continua, i.e. around the point of subjective equality (PSE), coinciding with reduced categorical perception. This was an unexpected report because, in terms of articulation /pae/, /tae/ and /kae/ are stop consonants with articulatory placement ranging from bilabial to alveolar to velar, respectively. One possible explanation for the perception of /kae/ where the stimuli is actually approximately 50% /pa/ and 50% /ta/ is that, taking into consideration the motor theory of speech (Rizzolatti and Craighero, 2004), such patients are incorrectly identifying the intended gestures of the vocal tract.

Unsurprisingly, this anecdotal data was only provided by patients and not in the healthy control population. It would be interesting to pursue this explanation further and compare speech perception specifically in terms of the motor theory of speech, in both the healthy and clinical populations.

4.3.6. Conclusions

In conclusion, through the analysis of categorization performances of right fronto-temporal stroke patients, this study shows that voice/gender processing depends critically on the right frontal cortex, likely the ventral part of the IFG. The specific functions of this frontal role are hypothesised to be the categorisation process, and possibly a more general role in social communication and vocal recognition. In contrast, left fronto-temporal patients (aphasic or not) tend to show associated performances (both impaired or not) for both voice/gender perception and speech perception, although dissociations are possible. Together, these results lend support to the hypothesis of bilateral processing of voice information with (i) an important role of the right frontal cortex in voice categorization and (ii) both common and dedicated mechanisms, in the left hemisphere, for talker and speech information processing, as phonetic and voice processing cannot be separated.

Finally this study holds possible implications for speech and language therapy in stroke patients. In right hemispheric stroke patients a voice perception impairment has been shown to be present in cases where speech perception is intact. Therefore, diagnosis of such an impairment may be useful in alerting patients to potential subsequent difficulties, for example in determining voice identification over the telephone. However, especially important, is that the results from this study are taken into consideration when assessing and treating *left* hemisphere damaged stroke patients. It is important to note that, as other studies have also shown (Turkeltaub et al., 2012; Turkeltaub and Coslett, 2010), perception impairments do not necessarily predict auditory comprehension deficits in people with aphasia. However, 9/10 patients had reduced perceptual distance (although only 4 dissociated), which suggests that voice/gender perception assessment may be beneficial for those with left hemispheric lesions who show impairments in speech perception, and thus suggests that it is not assumed that speech is always the underlying cause. Starting speech and language therapy at a low-level acoustic processing/ voice perception, could be a more appropriate and useful method in treatment of phoneme perception impairments than commencing directly with auditory comprehension tasks, as is current practice.

5. Chapter 5 – Crossed Aphasia

5.1. Introduction

Having introduced aphasia (Chapter 2) and then further investigated impaired speech and voice perception and its link to aphasia (Chapters 3 & 4), this chapter builds on these previous studies by looking at a relatively rare type of aphasia called 'Crossed aphasia'. The study aims to explore the behavioural and neurological presentation of a single case study of crossed aphasia, as well as investigating associated cognitive impairments through the single case study and a review of the literature. The research presented here further aims to look at the symptomology of crossed aphasia in the light of the cognitive-neurological models presented in Chapter 2.

In summary, the review of literature identified 4 main cognitive co-morbidities which are significantly associated with crossed aphasia. The case presented is of a gentleman with confirmed crossed aphasia with dyslexia and dysgraphia, in which the latter two cannot be fully explained by the current lesion and are probable developmental disorders (dyslexia/dysgraphia). Extensive longitudinal cognitive investigations and a series of advanced imaging techniques (structural and functional) were used to further investigate the cognitive and neuroanatomical basis of crossed aphasia and associated impairments (including voice/speech perception) in this patient.

Using the results from the literature review and the single case study, it is suggested that developmental disorders can be an underlying cause of partial right lateralisation shift of language processes, thereby supporting the theory that developmental disorders can be an underlying cause of crossed aphasia.

5.2. Study 3- Crossed aphasia and developmental disorders: a mini-review of the literature and investigation of a case study

5.2.1. Background

Article:

Jones, A.B., Bak, T., Bastin, M.E., Wardlaw, J.M., & Pernet, C. (2016). *Does crossed aphasia originate from developmental disorders? A Mini-review and case study*. Submitted for publication in Brain and Language.

Jones, A.B., Bak, T., Bastin, M.E., Wardlaw, J.M., & Pernet, C. (2016). *Does crossed aphasia originate from developmental disorders? A Mini-review and case study*. Currently online for peer-review at bioRxiv: doi:10.1101/039024.

Oral presentation:

Jones, A. *Crossed aphasia – a single case study*. Paper presented at the Aphasia Research Group, University College London; 2015 December 09, London, UK.

Jones, A. *Crossed aphasia case study*. Paper presented at The University of Edinburgh's Clinical Neuroscience seminar series; 2013 June 06, Edinburgh, UK.

Contributions by Anna Jones: conception and design of study and methodology, ethical submission, recruitment/consent, data collection and behavioural analysis, results interpretation, writing of and commenting on manuscript, writing and giving oral presentation.

In 1863, Paul Broca (1863) described a strict anatomo-functional connection between the handedness of a patient (i.e. right dominant motor control) and the hemispheric control over language functions. Twenty-eight years later, Oppenheim (1891) presented two cases of right-handed individuals who suffered right hemisphere lesions and subsequent aphasia, thus questioning the hypothesis of motor-language codominance. Following these observations, Bramwell (1899) proposed the term 'Crossed Aphasia' to denote aphasia caused by an ipsilateral lesion to the dominant hand. Wada and Rasmussen (1960) later showed that ~70% of left-handed patients also have left hemispheric dominance for language, thereby supporting Goodglass and Quadfasel's study (1954) earlier suggesting that crossed aphasia occurs in 70-80% of left-handed cases. This evidence showed that crossed aphasia in left-handed patients was actually the norm, and therefore the term 'Crossed Aphasia' (CA) became a synonymous term for 'Crossed Aphasia in Dextrals' (CAD). Specific criteria for a diagnosis of 'crossed aphasia' were first put in place by Brown and Wilson (1973) and reviewed by Habib et al.(1983) and Coppens and Robbey (1992). Mariën et al. (2004) conducted a review of the CA literature and concluded an algorithm of diagnostic criteria for vascular CA in adults widely in use today. This algorithm divides patients into unreliable, possible and reliable CA cases. Patients are only classed as reliable CA cases if they have 1. clear-cut evidence of language disorder, right-handedness and morphological integrity of the left hemisphere; as well as 2. absence of left-handedness in relatives and/or early brain damage/seizures in infancy (Figure 5-1).

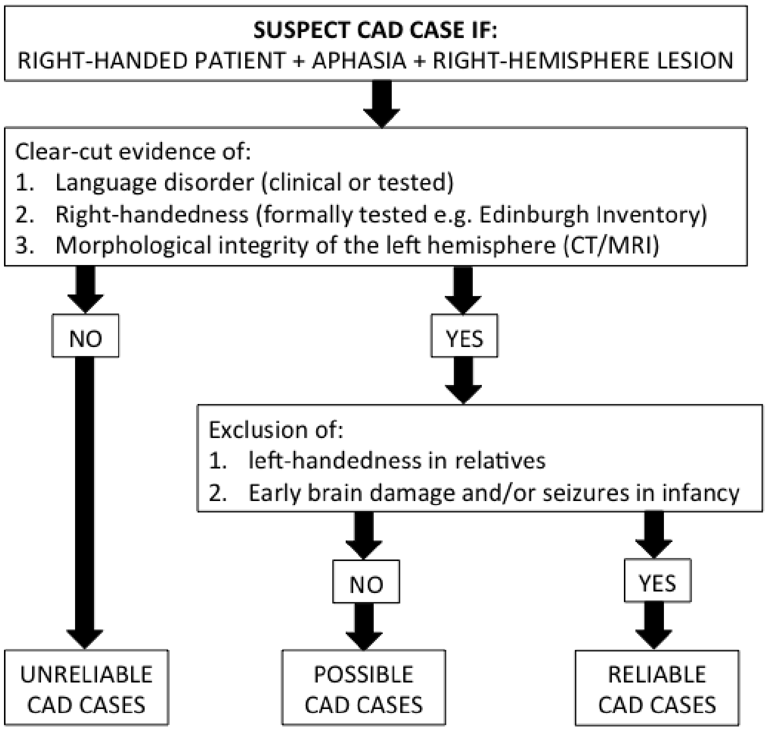

Figure 5-1: Algorithm of diagnostic criteria for vascular CAD in adults.

Redrawn from Mariën et al.'s review of Adult crossed Aphasia in Dextrals (2004).

To establish a clear-cut language disorder, and further understand the underlying neural substrates associated with CA, it is important to investigate co-morbidities (Marien et al., 2001). Co-morbidities are already widely acknowledged in the literature but are seldom investigated beyond the level of diagnosis. Dysgraphia for instance is frequently reported in dextral patients with aphasia from a left-hemisphere lesion (for example: Roeltgen and Heilman, 1984; Sinanovic et al., 2011; Tanridag and Kirshner, 1985), but there are only a few reports of patients with confirmed CA exhibiting dysgraphia (Assal et al., 1981b; Marien et al., 2001; Mastronardi et al., 1994) and no known reports of anomic crossed aphasia with persistent dysgraphia.

Specific lesion localisation, anatomical mapping and network disruption within CA has not been studied extensively either. Kim et al.'s review (2013) found that the regions most frequently involved in CA are the right lentiform nucleus, in particular the putamen, and basal ganglia. The right parahippocampal gyrus, claustrum, frontal lobe and precentral gyrus were also locations found to be involved in CA (Kim et al., 2013).

There are four main hypotheses for the possible causes underlying CA (Cappa et al., 1993): 1. dysfunction within the left hemisphere, either congential or acquired, causing a lateralisation shift (Bakar et al., 1996; Bhatnagar et al., 2006; Cappa et al., 1993). 2. Bilateral representation of language functions 3. Genetic basis (Alexander and Annett, 1996; Osmon et al., 1998) – i.e. the 'right-shift' (RS) theory (Annett, 1985). Cohen et al. (1993) also suggest a genetic underpinning but in relation to the co-existence of anomalous cerebral language dominance and situs inversus. 4. Diaschisis i.e. language areas in the left hemisphere are functionally depressed as a result of structural lesions in the right hemisphere. Therefore, a review of co-morbidities associated with CA was undertaken in order to test these hypotheses. In particular, if other deficits of known lateralized cognitive processes are associated to CA, this will allow the rejection of hypotheses 1 (only auditory and/or spoken language if right lateralized) and 2 (all language functions are bilateral) and further distinguish between hypotheses 3 and 4 (true right lateralization versus diachisis).

This study also details a new case of CA in a right-handed English-speaking male (CF) who suffered from an acute right middle cerebral artery infarct. His strengths and weaknesses are investigated using extensive cognitive neuropsychological assessments, as well advanced imaging methods (high resolution structural MRI, DTI and fMRI) to investigate lesion location, network disruption and language lateralization pattern. His deficits are then discussed in light of the results obtained in the review.

5.2.2. Methodology

Mini-review

A search of the literature was conducting using the following parameters in PubMed and ScienceDirect: ((crossed aphasi*[Title]) OR (crossed dysphasi*[Title]) OR ((right hemisphere stroke[Title]) AND (aphasi*[Title])) OR ((right hemisphere stroke[Title]) AND (dysphasi*[Title]))).

82 articles were retrieved, and using Marien et al.’s algorithm (2004) to select only reliable cases of CA, 23 articles were selected. A further 7 articles were retrieved from cited references, totalling 30 papers with 57 ‘reliable’ CA cases. After listing all co-morbidities, the 6 most common were selected and the number of cases was adjusted for the number of cases tested (e.g. a disorder might be present in all patients, but we don’t know about it because not tested). The adjustment was computed as the number of cases presenting with a given deficit * total number of CA cases/number of patients tested for that deficit. As there is no current epidemiological data available for the cognitive co-morbidities identified with crossed aphasia, a simple binomial model which assumes equal probability of occurrence for each condition was used (i.e. occurrence above 33%, corrected for multiple comparisons, upper bound of the 95% confidence intervals, is considered significant).

Among all co-morbidities, apraxia was split into “Central” (i.e. apraxia caused by impaired initiation e.g. constructional and ideomotor) and “Peripheral” (i.e. apraxia caused by impaired execution e.g. limb and oral) and only “central” apraxia was included since “peripheral” apraxia is not related to the language system. Visuo-spatial deficits were classified separately to visual-field neglect (a neuropsychological condition causing a deficit in attention to, and awareness of, one side of space) as they often occurred separately and/or were not both assessed. Visuo-spatial impairments observed across studies corresponded mainly to difficulties with visual organisation, spatial relations and position discrimination, constructive abilities, visual memory and visual scanning speed. Due to the heterogeneous nature of these impairments and the assessment procedures used in the studies included, visuo-spatial impairments were not included as one of the main cognitive co-morbidities in this review.

Case study

Ethics Statement

This study was approved by the NHS Lothian South East Scotland Research Ethics Committee 01 (Research Ethics Committee reference number: 11/SS/0055). Full information was provided to the participant, in an aphasia-friendly format (Appendix 7) and full informed consent was obtained from the participants (Appendix 8).

Diagnostic

To confirm the presence of crossed aphasia the integrity of the left hemisphere was investigated on CT (acute phase) and MRI (chronic phase – T1 3D IRP). Handedness was assessed using interview and the Edinburgh Handedness Inventory test (Oldfield, 1971). Aphasia was assessed using the Western Aphasia Battery (WAB) (Kertesz, 1982) at three separate time points: acute (1-6 weeks post stroke), sub-acute (2-6 months post stroke) and chronic (7-11 months post stroke).

Cognitive Assessment

Neuropsychological testing was conducted to assess CF’s language, speech/voice perception, memory, visuo-spatial skills, executive functioning, calculia (Ardila and

Rosselli, 2002; Leff et al., 2009) and retention of specific knowledge relate to his pre-morbid field of expertise (Graham et al., 1999a; Graham et al., 1999b; Omar et al., 2010; Robinson et al., 1999).

To explore the exact breakdown of CF's language, assessments were carried out using subtests of the Psycholinguistic Assessment of Language Processing in Aphasia (PALPA, Kay et al., 1992); an experimental battery of semantic assessments and an informal auditory phonological processing assessment (Supplementary Information 8). PALPA 1 (non-word minimal pair discrimination) and PALPA 2 (real word minimal pair discrimination) were used to assess phonological processing. PALPA 5 was used to assess auditory lexical decision-making. PALPA 8 (non-word repetition) and PALPA 13 (digit span) were used to assess auditory input and spoken output. PALPA 39 *(*words of varying letter length), PALPA 40 (words with varying imageability and frequency) and PALPA 45 (non words dictation) were used to assess auditory input. PALPA 8 and 36 (non word reading) was used to assess written input (reading) and spoken output. PALPA 54 (picture to written and spoken outputs) was used to assess object input and written/spoken output. Semantic assessments consisted of The Pyramids and Palmtrees Test (Howard and Patterson, 1992); Kissing and Dancing Test (KDT, Bak and Hodges, 2003); Tomato and Tuna Test (TTT, Faber et al., 2008); and Sound to Picture Matching Test (SPMT, Bozeat et al., 2000). All of these assessments were shortened, adapted and compiled for experimental purposes and programmed to run in E-prime 2.0 (Schneider et al., 2012) on a laptop computer. The raw score gained from each of these assessments was the number of errors made.

Speech and voice perception were assessed using an identical assessment to the experimental test used in Chapter 4 (Jones et al., 2015). This consisted of two alternative forced choice identification tasks: voice gender (male vs. female) and phoneme (/pae/ vs. /tae/) (both pitch-equalised). Between each task an interfering tone discrimination task was also performed, identical to that presented in Chapter 4 (Jones et al., 2015). Tasks were administered through Matlab (2014) on a Dell Latitude E5510 laptop, CF was asked to press keys on the keyboard (detailed instructions were presented for each task along with examples) and all sounds were presented binaurally through Vivanco 67 VSR headphones. Results were classified as ‘impaired’ or non-impaired’ for the voice gender and phoneme tasks, according to the criteria specified in Chapter 4 (Jones et al., 2015) and compared to data from right and left hemisphere damaged stroke patients, with and without an associated aphasia.

Memory was assessed testing verbal and non-verbal immediate recall, delayed recall, and processing speed [Doors subtest of the Doors and People Assessment, Brit Memory and Information Processing Battery, Rey-Osterrieth Complex Figure Test, Digit Span subtest of the Weschler Adult Intelligence Scale-IV, Weschler Memory Scales III] (Baddeley et al., 1994; Coughlan et al., 2007; Fastenau et al., 1999; Wechsler, 2008; Wechsler, 1999). Visuo-spatial skills were assessed using subsections of the Visual Object & Space Perception Battery (Lezak et al., 2004) as well as the ‘direct copy’ subsection from the Rey-Osterrieth Complex Figure task (Fastenau et al., 1999). Executive Functioning was investigated using the matrix reasoning subtest from the WAIS IV (Wechsler, 2008) and the Key Search, Temporal Judgement, and Modified 6 elements subtests from the Behavioural Assessment of Dysexecutive Function battery (Wilson et al., 1996). Acalculia was informally tested using the following mathematical tasks: addition, subtraction and deciding which is the larger/smaller out of two numbers. Premorbid semantic knowledge was investigated using an informal questionnaire (Appendix 9).

Standardised assessment results were compared against published normative data and performance descriptors (impaired/unimpaired) applied. If published normative data was unavailable results were compared against published control data. An impaired performance descriptor was applied to all scores below 100% on informal assessments without published control data (apart from the voice/speech perception tasks). Results were mapped onto an adapted and extended cognitive-neuropsychological model of single word processing (Figure 5-5) with the aim of identifying the nature of the underlying impairments. Assessments were matched to the model according to the different processing components used within the task. Performance descriptors, error analysis and convergent evidence from different assessments were then used to identify if a processing component within the module was intact or impaired (Whitworth et al., 2014).

Cognitive-neuropsychological models of single word processing (comprehension and expression) are static models, allowing an individual's performance to be mapped at a single time-point. The model used in this case study is an extended version of Whitworth's (2014) adaptation of the Ellis (1988) model of language processing for single words. The model further includes a phonological input buffer (Laganaro and Alario, 2006; Martin et al., 1999; Nickels et al., 1997), phonological output buffer and orthographic input buffer (Kay et al., 1992), as well as including dual processing routes (Ellis et al., 1988). Spoken language is represented on the left, written language on the right. Five main processing routes are shown: speech perception (top left); speech production (bottom left); reading (top right); writing (bottom right); and visual object input (top centre). The semantic system (along with the object concept system) is the main commonality to all routes (centre, red). Boxes represent independent modules; arrows represent flow of information (in the direction of the arrow) between the specified processor/module; blue rectangles represent input and output modalities; dashed lines represent information flow that is bypassing modules; dashed boxes represent modules which are involved in this 'bypass' process; "access" represents flows of information that can become impaired.

Imaging

In addition to the CT scan performed at the acute time point, high-resolution structural, DTI and fMRI data were obtained at the chronic phase (75 weeks post-stroke onset) using a GE Signa HDxt 1.5 T scanner at the Brain Research Imaging Centre (http://www.bric.ed.ac.uk/), University of Edinburgh, UK. Structural imaging consisted of the following sequences: (i) a T1-weighted volume (3D IRP - 180 slices, 2 mm thick coronal slices, 1.3 x 1.3 mm in-plane resolution with a 256 mm FOV); (ii) a T2-weighted volume (FSE - 72 slices, 2 mm thick axial slices thickness, 1 x 1 mm in-plane resolution with a 256 mm FOV), and (iii) a FLAIR-weighted volume (FSE –

40 slices, 4mm thick axial slices, 1 x 1.3 mm in-plane resolution with a 256 mm FOV); and (iv) Diffusion Weighted Images (DTI - 72 slices, 2 mm thick axial slices, 2 x 2 mm in-plane resolution with a 256 mm FOV). The DTI examination consisted of 7x T2-weighted (b = 0 s mm-2) and sets of diffusion-weighted (b = 1000 s mm-2) single-shot spin-echo echo-planar imaging (EPI) volumes acquired with diffusion gradients applied in 64 non-collinear directions. Volumes were acquired in the axial plane, with a FOV of 256 × 256 mm, contiguous slice locations, and image matrix and slice thickness designed to give 2 mm isotropic voxels. The repetition and echo time for each EP volume were 16.5 s and 98 ms respectively. fMRI was also performed to map language areas. The auditory cortex and Wernicke area were mapped using a word repetition task whilst Broca's area was mapped using a verb generation task (Gorgolewski et al., 2013a), and the Visual Word Form Area was mapped using a visual detection task (see details below). The 'temporal voice areas' (TVA) of the auditory cortex were mapped using a voice localiser task (Pernet et al., 2015a). The fMRI data was acquiring as follows: a) *Word repetition task* and b) *Verb generation task* - (EPI – 30 slices, 4mm thick axial slices, 4 x 4 mm in-plane resolution with a 256 mm FOV); c) *Passive visual word form task* - (EPI – 27 slices, 4 mm thick axial slices, 4 x 4 mm in-plane resolution with a 256 mm FOV); d) *Voice localiser task* – (EPI - 27 slices, 4 mm thick axial slices, 4 x 4 mm in-plane resolution with a 256 mm FOV).

For the word repetition task (a) CF was asked to repeat words aloud as per the instructions: "When you hear the word, repeat it immediately". After 4 trial practices, stimuli were presented as a block design (30 sec on/off, repeated 6 times) and a sparse sampling acquisition. Nouns (selected randomly) were presented via headphones during the silent periods followed by the visual presentation of a question mark prompting repetition out loud. During the rest periods a blank screen was presented.

For the Verb generation task (b), CF was asked to think of a verb complementing a noun presented in written form. Instructions were as follows: "When a word appears it will be a noun. Think of what you can do with it and then imagine saying 'With that I can …' or 'That I can …' " Again, after 8 practice trials, stimuli were presented following a block design (30 sec on/off, repeated 6 times). During each activation period, 10 nouns were presented in written form for 1 sec each, followed by a fixation cross during which time CF generated a covert response. The nouns were selected randomly from a set of 70 (Gorgolewski et al., 2013a). Each rest period was similar to the activation periods; however a scrambled form of a visual picture of the noun replaced the written word and nothing was expected from the patient.

For the visual word form area mapping task (c), CF viewed pictures and was asked to respond by pressing a button when a picture previously seen was repeated (occurred 1 or 2 times in 3 out of 8 blocks per stimulus category). He received these instructions: "You will see a number of images, please press a button every time you see an image you have previously seen." A block design was used with activation blocks of 16 seconds, corresponding to faces, objects, words or checkerboards. The task lasted a total of 10 minutes and 56 seconds with 8 blocks per category. During the activation periods 24 stimuli were presented for 330ms and then 330ms off. Words presented were high frequency English nouns (3-7 letters) (taken from http://www.esldesk.com/esl-quizzes/frequently-used-english-words/words.htm) (Cohen et al., 2000a; Dehaene and Cohen, 2011; Price and Devlin, 2011). Object pictures were taken from the Amsterdam Library of Object Images (Geusebroek et al., 2005) and from the Department of Cognitive, Linguistic & Psychological Sciences, Brown University (http://titan.cog.brown.edu:8080/TarrLab). Face pictures were taken from 'Labelled Faces in the Wild' (http://vis-www.cs.umass.edu/lfw/index.html). All pictures were changed to gray and root-mean squared (RMS) contrast. During the rest periods a fixation cross was presented.

For the voice localiser task (d), CF was instructed to passively listen to vocal and non-vocal sounds. It is a fixed block design with forty x 8 second long blocks or vocal (50%) or non-vocal (50%) sounds. Stimuli were taken from http://vnl.psy.gla.ac.uk/resources.php (Pernet, McAleer et al. 2015). A sparce-sampling design was used, to avoid stimulus masking by scanner noise. The task lasts for 10 mins 20 sec in total.

Imaging data

1. Structural processing

Structural images were denoised using SUSAN (Smith and Brady, 1997) from the FSL (Jenkinson et al., 2012) library and the T2-weighted and FLAIR images were co-registered to the T1-weighted images using SPM12. Images were centred along the anterior commissure and segmented images produced using mulitspectral segmentation from SPM12. The gray matter image was then thresholded at 0.2 and smoothed with a 4mm isotropic Gaussian kernel to create a gray matter mask.

2. fMRI data analysis

For each fMRI task SPM12 was used. Data were first slice-time corrected (amount of correction varied according to specific task parameters, but in all cases the data was temporally aligned to the middle temporal slice), then realigned to the 1st image of each session and then to the mean EPI (SPM12 default parameters) and finally smoothed at 6mm isotropic Gaussian kernel. The T1 image (following the structural processing above) was then co-registered onto the mean EPI and the transformation parameters applied to the gray matter mask created earlier. The General Linear Model (Friston et al., 1994) was used to estimate the BOLD signal response for each task separately with parameter estimates restricted to the gray matter mask. For each task, one regressor per condition was used (1 regressor for activation blocks in the word repetition; 1 regressor for activation blocks in the verb generation task; 4 regressors for the blocks of faces/ objects/ words/ checkerboards in the passive visual work form task; 2 regressors for vocal sounds/ non-vocal sounds in the voice localiser task) as well as motion parameters and motion outlier censoring(Siegel et al., 2014). Adaptive thresholding (Gorgolewski et al., 2012) was used to obtain the single subject statistical maps for each task.

Hemispheric lateralisation of the whole brain and the temporal lobe for the Word repetition task (a) and the Verb generation task (b) was also assessed, based on number of activated voxels, using the LI-tool (Wilke and Lidzba, 2007) in SPM12, producing lateralisation indices (LI) in each task (LI>0 signifies left hemispheric lateralisation, LI<0 signifies right hemispheric lateralisation). To allow for comparison, fMRI data from 10 healthy control participants (median age 52.5 years, 3 left-handed and 7 right-handed) (Gorgolewski et al., 2013b) who had carried out the same two tasks was also analysed for hemispheric lateralisation of the whole brain and temporal lobe, using the LI-tool (Wilke and Lidzba, 2007).

3. Tract-based Spatial Statistics

All DTI data were converted from DICOM (http://dicom.nema.org) to NIfTI-1 (http://nifti.-nimh.nih.gov/nifti-1) format using the TractoR package for fibre tracking analysis (http://www.tractor-mri.org.uk). FSL tools (http://www.fmrib.ox.ac.uk/fsl) were then used to extract the brain, remove bulk motion and eddy current induced distortions by registering all subsequent volumes to the first T2-weighted EPI volume, estimate the water diffusion tensor and calculate parametric maps of MD and FA from its eigenvalues using DTIFIT.

Following protocols described in detail by ENIGMA (Enhancing Neuro Imaging Genetics Through Meta Analysis; http://enigma.ini.usc.edu/protocols/dti-protocols/#eDTI), differences in CF’s MD values in language tracts (left and right internal capsule, inferior and superior longitudinal fasciculi and splenium of corpus callosum) compared with 5 aged matched (63.3 ± 1.4 years) healthy controls were assessed in regions-of-interest (ROI) extracted from white matter skeletons produced using Tract-based Spatial Statistics (TBSS; http://www.fmrib.ox.ac.uk/fsl). First, all FA volumes were linearly and non-linearly registered to the standard FMRIB58_FA volume. Second, a white matter skeleton was created from the mean of all registered FA volumes. This was achieved by searching for maximum FA values in directions perpendicular to the local tract direction in the mean FA volume. An FA threshold of 0.25 was applied to the skeleton to exclude predominantly non-white matter voxels. Third, for each subject’s FA volume, the maximum voxel perpendicular to the local skeleton direction was projected onto the skeleton. This resulted in one FA skeleton volume per subject corresponding to centres of white matter structures. Average MD values were then obtained for the six subjects from skeletal projections in language tract ROI defined using FSL's JHU white matter atlas. Analysis of DTI data was performed by Dr. Mark Bastin (MRI physicist).

5.2.3. Results

Mini-review

Using Marien et al.'s algorithm (2004) 57 'reliable' CA cases were identified, 52 as a result of stroke and 5 as a result of a tumour (Alexander and Annett, 1996; April and Han, 1980; Assal et al., 1981a; Bakar et al., 1996; Bartha et al., 2004; Bhatnagar et al., 2006; Bhatnagar et al., 2011; Cappa et al., 1993; Cohen et al., 1993; De Witte et al., 2008; Denes and Caviezel, 1981; Faglia and Vignolo, 1990; Giovagnoli, 1993; Ha et al., 2012; Haaland and Miranda, 1982; Habib et al., 1983; Henderson, 1983; Ishizaki et al., 2012; Kim et al., 2013; Lessa Mansur et al., 2006; Marien et al., 2001; Marshall and Halligan, 1992; Mastronardi et al., 1994; Osmon et al., 1998; Paghera et al., 2003; Paparounas et al., 2002; Patidar et al., 2013; Rey et al., 1994; Stefanis et al., 1997). Across those studies, six main co-morbidities were found: Central Apraxia, Dysgraphia, Hemi-neglect, Acalculia, Attentional deficits, Memory deficits (Figure 5-2). Under the hypothesis they all have equal chance to co-occur (chance level: 6 to 33%), central apraxia, dysgraphia, left visual field neglect and acalculia were found to significantly co-occur with CA (Table 5-1). (See Appendix 10 for a detailed review of studies involved).

Cognitive co-morbidity	**Number of cases tested**	**Number of cases identified**	**Adjusted number of cases**
Central Apraxia	31	21	39
Dysgraphia	36	22	35
Left VF neglect	33	22	38
Acalculia	18	13	42
Attention	14	4	15
Memory	15	5	19

Table 5-1: Breakdown of cases detailed in literature review according to cognitive co-morbidities, showing the total number of cases assessed and associated number of cases identified.

Adjusted number of cases is the estimated number of cases out of the unique cases of confirmed CA. Italics show co-morbidities significantly likely to co-occur with CA (i.e. not at chance level). Apraxia was split into "Central" (i.e. apraxia caused by impaired initiation e.g. constructional and ideomotor) and "Peripheral" (i.e. apraxia caused by impaired execution e.g. limb and oral) apraxia. Only "central" apraxia was included as a main cognitive co-morbidity because "peripheral" apraxia is not related to the language system. 'Other visuo-spatial difficulties' identified were deemed too broad to include.

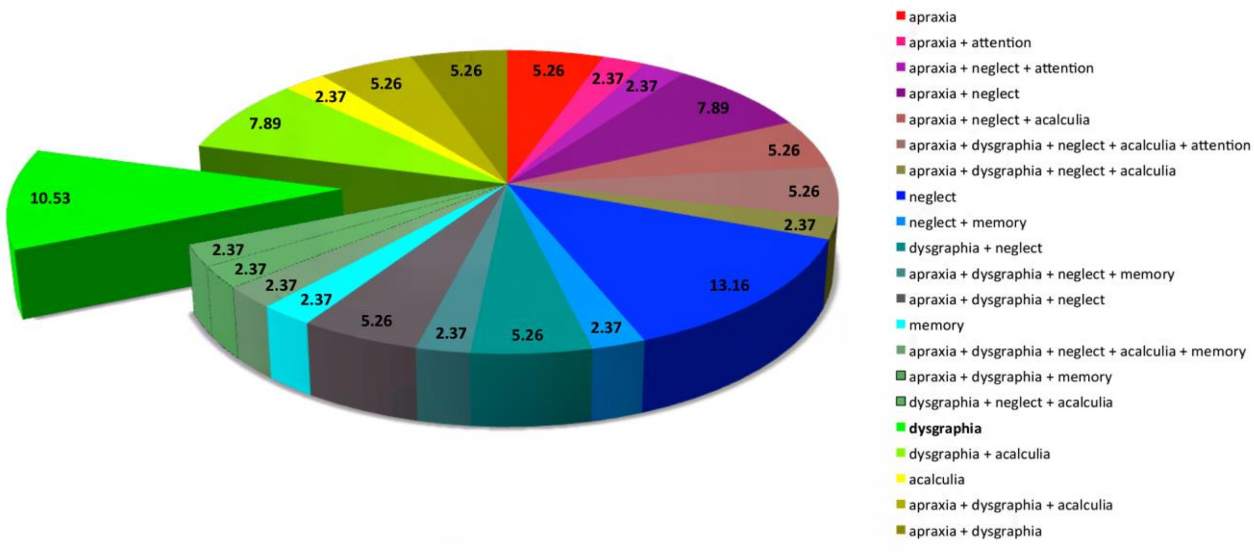

Figure 5-2: Pie chart showing occurrence of the main cognitive co-morbidities present alongside confirmed cases of Crossed Aphasia in the literature

The 6 main co-morbidities identified are: central apraxia (apraxia); difficulties of attention (attention); left visual field neglect (neglect); acalculia (acalculia); dysgraphia (dysgraphia); and memory impairments (memory). Percentages correspond to the number of times one or more cognitive impairment co-occurred (total number of classes identified =21) / total number of confirmed CA cases with at least one of the main co-morbidities present (n=38). The exploded pie segment (dysgraphia) is the class of co-morbidities corresponding with CF's symptomatology at the chronic stage.

Aphasia and apraxia are two independent conditions, but they are often associated. The language and praxis systems share a number of functional features (i.e. sensory-motor integration and symbolic representation) and rely on common anatomical structures involving the frontal cortex and basal ganglia (Gross and Grossman, 2008; Kobayashi and Ugawa, 2013). Double dissociations have however been found between aphasia and apraxia (Papagno et al., 1993), suggesting that the two networks do not completely overlap. Praxis relies on a large-scale network involving frontal-parietal-regions and basal ganglia. Central apraxia is commonly associated with right hemispheric damage, and is most likely caused by a higher-order visuospatial processing deficit in patients with parietal (intraparietal sulcus) lesions, or by impairments to organisation and planning in patients with frontal (middle frontal gyus - MFG) lobe damage (Haaland et al., 2000). As both the intraparietal sulcus and the MFG are involved in language (syntax (Carreiras et al., 2010), dorsal language pathway (Herman et al., 2013) and phoneme detection (Simon et al., 2002), lexical-syntactic retrieval (Acheson and Hagoort, 2013), respectively), it stands to reason that central apraxia significantly co-occurs with CA, if we consider that only language is inversely lateralized in those individuals.

Dysgraphia impairments observed across studies corresponded to misspelling of words; neologisms, paraphasias and jargon; perseveration of letters or words; semantic errors; syntactical errors; spatial problems; morphological errors; word or letter omissions, substitutions, additions and deletions; and motor difficulties. Central writing processes are defined as 'the retrieval of abstract orthographic word-forms, via orthographic lexicon or phoneme-to-grapheme conversion mechanisms, and their temporary storage in the graphemic buffer' (Planton et al., 2013) (p. 2773). Note that motor and linguistic impairments involved in writing are often associated with other abilities (i.e. praxis, literacy etc.) and thus not included in this definition. For the purposes of this study only those impairments involving 'writing specific' processes (i.e. orthographic coding - not including syntax, semantics, spatial or motor difficulties) were classified as 'dysgraphic'. Two out of the three major anatomical regions involved in the writing system are also damaged in some aphasias: the superior frontal gyrus (SFG) involved in the recollection of grapheme representations and the supramarginal gyrus (SMG) involved in phoneme-to-grapheme conversion (Planton et al., 2013). Under the constrictive criteria discussed above, dysgraphia significantly co-occurs with crossed aphasia, which is expected if one assumes that all language areas (i.e. not just perisylvian ones) are right lateralized.

Unilateral visual-field neglect is a common neurological presentation predominantly following damage to the right ventral fronto-parietal cortex i.e. a proposed distributed ventral attention network involving right frontal, temporal and parietal cortex (Corbetta and Shulman, 2011), which in turn disrupts the dorsal attention network. Importantly the dorsal attention network can be impaired (causing neglect symptoms) by damage to a variety of right hemispheric ventral fronto-parietal regions (Corbetta and Shulman, 2011). The data presented here showing that left visual-field neglect is significantly associated with CA are supportive of both the right hemisphere dominance of the visual attention system and also the wide variety of fronto-parietal regions corresponding with neglect. It has been postulated that these regions vary in terms of criticality to the overall attention process and network, with the more posterior regions more crucially involved (Callejas et al., 2014), specifically the inferior parietal lobule, temporo-parietal junction, the superior temporal lobe and the angular gyrus (Gillebert et al., 2011; Karnath et al., 2004; Karnath and Rorden, 2012; Mort et al., 2003). In cases of crossed aphasia, left visual field neglect commonly occurs alongside central apraxia which again is expected if only language is inversely lateralized in those individuals.

Acalculia significantly co-occurs alongside crossed aphasia with the highest prevalence overall. Numerical cognition consists of a fronto-parietal network involving intra-parietal and pre-frontal areas (Moeller et al., 2015). The triple-code model of numerical processing (Dehaene et al., 2003) details three circuits (visual system encoding Arabic numbers; quantity system encoding analogical-semantic representations of size and distance relations; verbal system encoding numerals lexically, phonologically and syntactically) co-existing in the parietal lobe, specifically in the bilateral superior parietal gyrus, bilateral intra-parietal gyrus and the left angular gyrus. The frontal section of the pathway involves the pre-frontal cortex, in particular the inferior, medial and superior frontal gyri (Simon et al., 2002). Fronto-parietal association fibres (superior longitudinal fasciculus dorsally and external capsule ventrally) are also involved in numerical cognition (Moeller et al., 2015). Therefore, lesions in a number of parietal regions can be attributed to both language and numeracy impairments, for example the intraparietal sulcus is involved in number processing/arithmetic calculations (Seghier et al., 2008) as well as numerous components of language processing [syntax processing (Carreiras et al., 2010), phoneme detection (Simon et al., 2002) and the dorsal language pathway (Herman et al., 2013)]. The significant co-occurrence of acalculia is expected as a consequence of the disruption between the quantity and the verbal systems, only this one being right rather than left lateralized.

Case study

CF was 66 year-old monolingual, right-handed English speaker male laterality quotient (LQ) = +100, Decline R.10 (Oldfield, 1971)) admitted to hospital with a left sided weakness (including a left sided facial droop), confusion, apraxia and aphasia. A clinical CT scan showed an acute right middle cerebral artery (MCA) infarct. After 5 days in an acute ward, he was transferred to a stroke rehabilitation unit where he received speech and language therapy (SLT) for 1 month. He was then discharged home and immediately received SLT weekly for the following 11 months. All acute motor impairments resolved prior to transfer to inpatient rehabilitation; however, CF was left with a mixed communication impairment (aphasia and dysgraphia). CF was high functioning in all aspects of his life prior to his stroke. CF noted that he had difficulties as a child with written spelling and had suspected developmental dysgraphia. No other communication or visual problems were present before CF's admission for this episode. CF had a bilateral, symmetrical sensorineural hearing loss, corrected by bilateral hearing aids. At the time of testing, he showed mild low mood but did not have clinical depression (Depression Intensity Scale Circles, (DISCS, Turner-Stokes et al., 2005) and Visual Analogue Self-Esteem Scale, (VASES, Brumfitt and Sheeran, 1999)).

Lesion location

A CT scan (acute phase), confirmed by an MRI scan (chronic phase), showed complete integrity of the left hemisphere and revealed in the right hemisphere: (i) low attenuation and loss of grey/white matter differentiation, with mild swelling, in the insula, internal capsule, frontal operculum, and part of the inferior frontal gyrus and the mid frontal gyrus, (ii) a loss of basal ganglia definition, (iii) a mild degree of mass effect associated with the ischaemic lesion, and (iv) hyperdense MCA at the level of bifurcation and proximal M2 branches, signifying an intravascular thrombus as the cause of the infarct. Examination of fractional anisotrophy (FA) maps obtained from diffusion tensor MRI (DTI) also suggests alterations in fibre structure around the internal capsule and arcuate branch of the superior frontal fasciculus. A statistical comparison of mean diffusivity (MD) and FA values of CFs' language tracts against 5 aged matched healthy controls not only confirmed reductions in structural integrity of the anterior, posterior and retro-splenial limb of the internal capsule and of the superior longitudinal fasciculus (Figure 5-3) but also defects in the right inferior fronto-occipital fasciculus and bilateral uncinate fasiculi (Table 5-2). For these last two tracks, both left and right MD values were much larger, and FA values much smaller, than in the controls.

TRACT	Volunteers		CF	
	MD (×10-6 mm2/s)	Laterality Index	MD	Laterality Index
Splenium	572 (28)		629	
Right anterior limb of internal capsule	602 (29)		682	
Left anterior limb of internal capsule	576 (37)	-0.02 (0.02)	621	-0.04
Right posterior limb of internal capsule	565 (24)		645	
Left posterior limb of internal capsule	576 (36)	0.009 (0.01)	590	-0.04
Right retrolenticular limb of the internal capsule	631 (28)		738	
Left retrolenticular limb of the internal capsule	638 (17)	0.006 (0.01)	719	-0.01
Right superior longitudinal fasciculus	611 (26)		740	
Left superior longitudinal fasciculus	626 (38)	0.01 (0.01)	690	-0.03
Right inferior fronto-occipital fasciculus	623 (31)		733	
Left inferior fronto-occipital fasciculus	676 (42)	0.04 (0.03)	721	-0.008
Right Uncinate fasciculus	728 (57)		1047	
Left Uncinate fasciculus	707 (36)	-0.01 (0.04)	1027	-0.009

Table 5-2: Average and standard deviations of Mean diffusivity (MD) values.

Table shows MD for 5 healthy controls aged 63.3 (S.D.=1.4) with corresponding Laterality Index (Lebel and Beaulieu, 2009) compared to CF (case study).

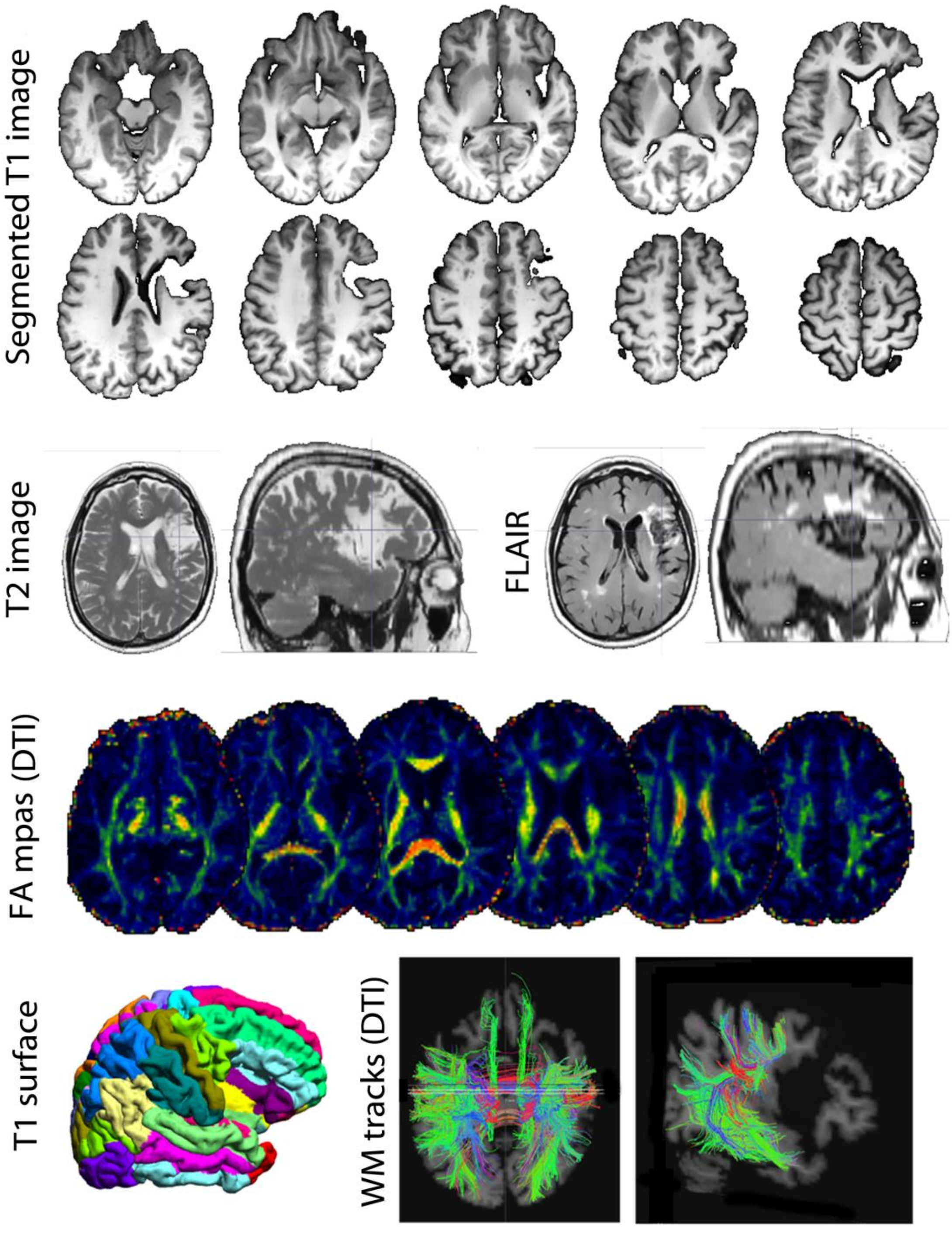

Figure 5-3: Structural imaging

At the top the segmented T1 image is shown, along with the corresponding T2 and FLAIR images, illustrating the location and size of the lesion. The bottom part of the figure show reconstructed images from the T1 (pial surface) and from the DTI data (Fractional Anistropy (FA) maps and White Matter (WM) tracks).

Language

Language was assessed longitudinally using The Western Aphasia Battery (WAB) (Kertesz, 1982): acute (1-6 weeks post stroke), sub-acute (2-6 months post stroke) and chronic (7-11 months post stroke). At the **acute** time point and sub-acute time points CF was classified as 'anomic' (Aphasia Quotient of 79.9 and 87.8 respectively). At the **chronic** time point, CF obtained an Aphasia Quotient of 96.4 and was no longer classified as aphasic. Residual deficits were mild word-finding difficulties (fluency score = 9/10) and written output difficulties (written score = 91/100). His written output was characterised by syntactical errors, occasional phonemic orthographic output errors and cognitive demand errors (Figure 5-5). Full breakdown of results in Appendix 11.

To explore the exact breakdown of CF's language selected subtests of the Psycholinguistic Assessment of Language Processing in Aphasia (PALPA, Kay et al., 1992), an experimental battery of semantic assessments and an informal auditory phonological processing assessment were carried out and mapped onto an adapted cognitive-neuropsychological model (Figure 5-5, Figure 5-8). See Appendix 14 for details of language assessments carried out and results.

<u>Phonological processing</u>, assessed using an informal auditory discrimination task at the acute/ sub-acute phase, showed an impaired distinction between /m/ and /n/; /k/ and /g/; /s/ and /z/. CF was, however, able to distinguish the following phoneme pairs: /p/ and /b/; /t/ and /d/; /ɵ/ and /ɸ/;/ʃ/ and /ʒ/;/ɾ/ and /l/. At the word level (assessed at the chronic phase), phoneme perception was also impaired with scores on PALPA 1 (non-word minimal pairs discrimination) (Kay et al., 1992) showing deficits for both 'same' and 'different' items, and scores on the PALPA 2 (real word minimal pairs discrimination) (Kay et al., 1992) showing deficits for the 'same' items but not for 'different' items. Together these results suggest that CF has impairment in auditory phonological analysis, which did not resolve between the acute and chronic phase. Assessment of <u>auditory lexical decision</u> using PALPA 5 (Kay et al., 1992) at the chronic phase, showed no impairment with real word decisions but outside of the

normal range of performance for non-words, indicating a breakdown in the phonological input lexicon. Auditory input and spoken output were assessed using PALPA 8 (non-word repetition) and PALPA 13 (digit span) (Kay et al., 1992) at the chronic phase. CF showed an overall ability to repeat non-words with a few errors when increasing syllable length. His auditory digit span was unimpaired for both digit repetition and digit matching, with no length effects; which indicates that non-words repetition errors were caused by phonological output buffer impairment. Auditory Comprehension was assessed using subsections of the WAB (Western Aphasia Battery) (Kertesz, 1982) at both the acute and chronic phases. Comprehension of single words and short phrases were within normal range at both stages. Comprehension of sentences was impaired at the acute phase, but was no longer impaired by the chronic stage.

Functional MRI revealed a right lateralized pattern of activation for the word repetition and verb generation tasks (based on bootstrap lateralization curves - independently of the statistical threshold used), that was even stronger than the one observed for the 3 left handed control subjects tested (controls tested: right-handed n=7; left-handed n=3 (Gorgolewski et al., 2012) (Figure 5-4). The word repetition task showed bilateral activations over the primary and secondary auditory cortices, extending posteriorly to Wernicke's area, similar to single subjects' control data (Gorgolewski et al., 2012). The verb generation tasks elicited activations in the right mid frontal gyrus and right posterior temporal cortex, almost mirroring the left frontal and temporal activations observed in controls (Gorgolewski et al., 2012).

Written input (reading) and spoken output were assessed by the WAB (word reading) (Kertesz, 1982) and both PALPA 8 and 36 (non-word reading) (Kay et al., 1992), at the chronic stage. Performances on the WAB were within normal range, indicating an intact lexical reading route. In contrast, non-word reading was impaired, with a length effect whereby non-words containing more than 2 syllables/ 4 letters were not read. This suggests an inability to hold long non-words in the phonological output buffer and/or the orthographic-phonological conversion module.

The Visual Word Form Area was mapped using a passive localizer contrasting high frequency words to checkerboards (Cohen et al., 2000a; Dehaene and Cohen, 2011; Price and Devlin, 2011). Results show right lateralized activations (LI -0.35), in particular over the homologue of the Visual Word Form Area (Cohen et al., 2000b). During the same localizer, the FFA was mapped contrasting faces with checkerboard, showing co-localization in the right hemisphere (LI = -0.2; Figure 5-4).

Object input and written/spoken output were tested at the chronic stage using PALPA 54 (picture to written and spoken outputs) (Kay et al., 1992). CF did not have any difficulties with single word expression from pictorial inputs although he did make a few semantically related errors. Object input and semantics: CF was unimpaired in all of the semantic processing (noun, verb, syntagmatic and sounds) assessments, showing no central semantic system impairment.

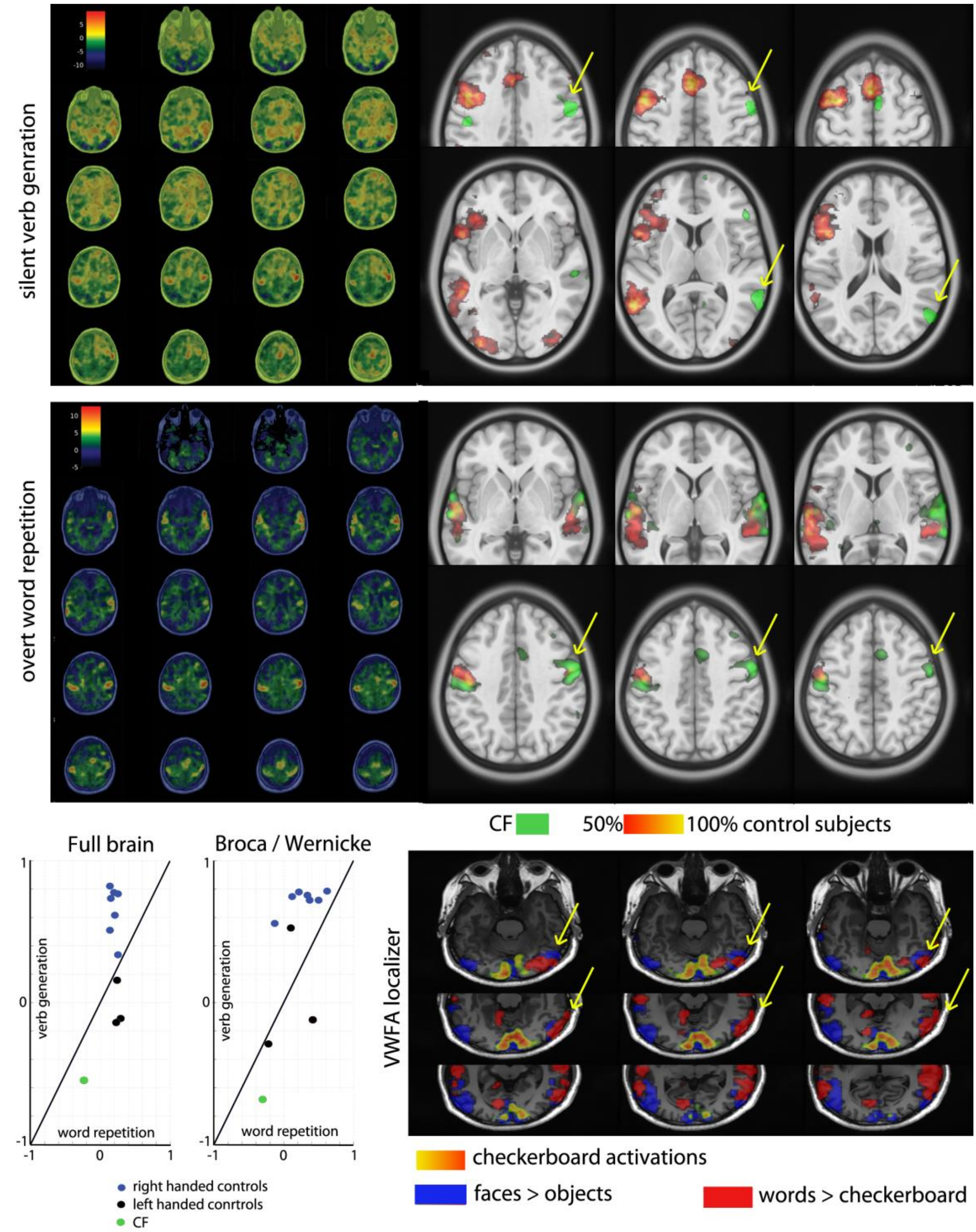

Figure 5-4: Functional Imaging results.

On the left side (top and middle rows) are the unthresholded fMRI maps of CF for the verb generation task (visual input) and the overt word repetition task (auditory input). At the bottom (left) are plots showing the lateralization index obtained in each task for CF compared to a control group of 10 subjects. The overlap of the single subject maps, projected into the standard space, can be seen on the right side (top and middle rows). The bottom right hand side shows the thresholded contrast maps for (left-right) checkerboard (V1), faces (FFA) and words (VWFA).

Dysgraphia

At the **acute** time point CF showed severely impaired written output on The Western Aphasia Battery (WAB) (Kertesz, 1982) (written score = 44.5/100). His written output improved at the sub-acute time point (written score = 66.5/100), although still showed the same pattern of errors as in the acute phase. At the chronic time point CF's written output was within normal limits, but was characterised by some syntactical errors, occasional phonemic orthographic output errors and cognitive demand errors (Figure 5-5, Figure 5-8).

Written output was further examined at both the acute and chronic stages using PALPA 39 (words of varying letter length), PALPA 40 (words with varying imageability and frequency) and PALPA 45 (non-words dictation) (Kay et al., 1992). At the acute stage, a severe dysgraphia was observed with an effect of word length, imageability and frequency and an inability to spell non-words. At the chronic stage, performances were within or close to normal for word spelling but still showed a mild letter length effect and imageability effect, and no frequency effect. Despite performance recovery for words, non-word spelling was still severely impaired, indicative of a dysgraphia with a breakdown in phoneme/grapheme correspondence and at the level of the graphemic output buffer. An important observation is the marked difference between CF's spoken spelling and his written spelling at both time points - CF tended to spell the words dictated to him aloud in the correct form, at the same time as writing them down in an incorrect form. This dissociation clearly shows that CF's written output system is selectively impaired, with the spoken output system remaining intact, supporting impairment in phoneme/grapheme correspondence.

Interestingly, all three major areas involved in the 'writing system' (SFG, SMG and intra-parietal sulcus (Planton et al., 2013)) were physically intact. The lesion affects frontal areas inferior (lateral IFG) to the 'writing system' (SFG (Planton et al., 2013)). The IFG and the SMG are both involved in grapheme/phoneme correspondence (Mei et al., 2014) and lesions of the fibers linking these regions could explain some of the observed deficits.

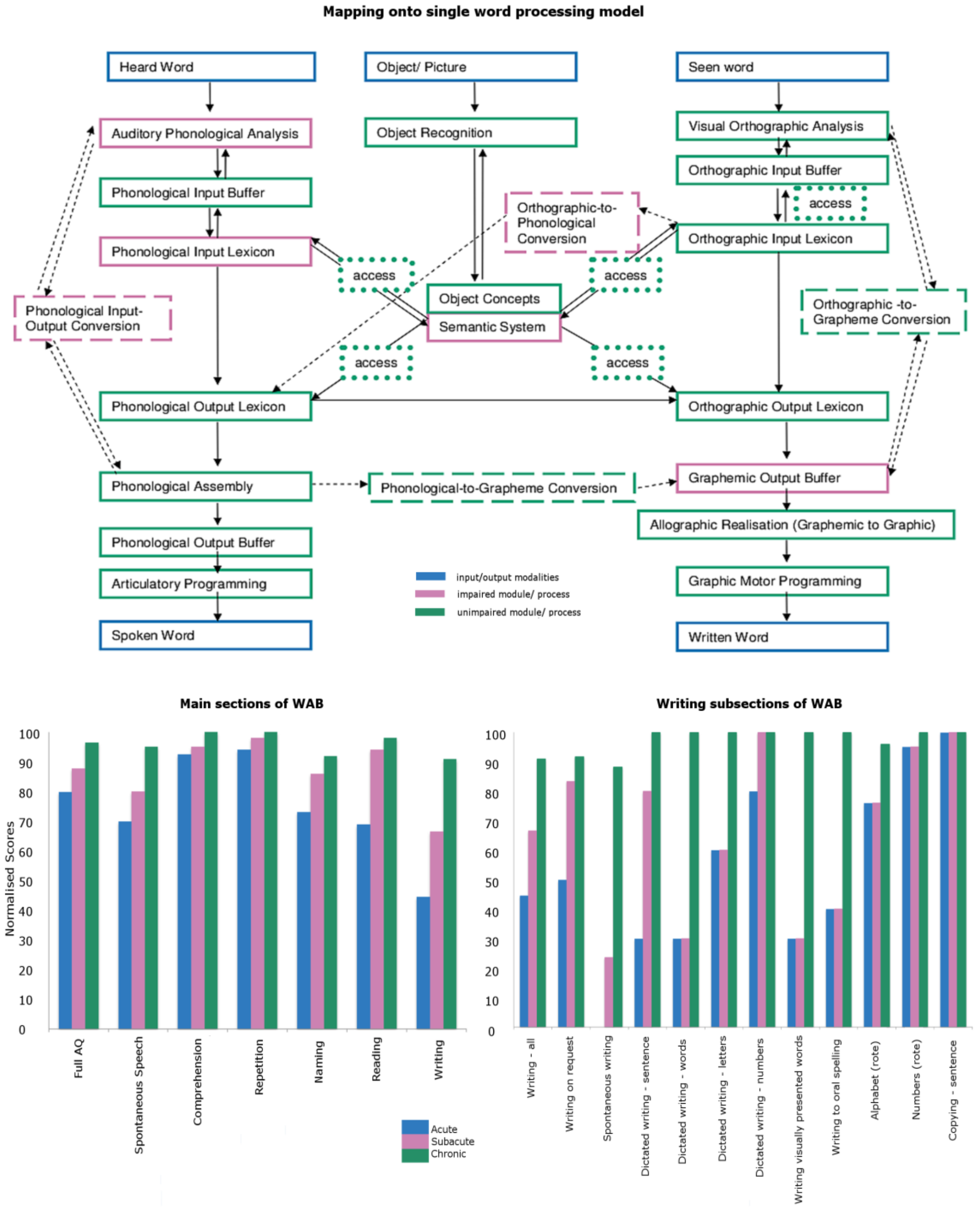

Figure 5-5: Behavioural scores from language assessments (simple method of normalising values to the sum of 100 was applied) and mapping onto single word processing model.

Top: CF's breakdown at chronic time-point mapped onto a cognitive neuropsychological single word processing model (adapted from Whitworth (Whitworth et al., 2014) and Ellis (2004)). Blue: input and output modalities; Green: unimpaired module (solid/ dashed box)/ process (dotted box); Pink: impaired module (solid/dashed box)/ process (dotted box).

Bottom: Bar-graphs showing CF's normalised scores on the Western Aphasia Battery at acute, sub-acute and chronic time, Left: main WAB sections; right: writing subsections.

Voice perception

Voice-sensitive areas were also mapped at the chronic stage using functional MRI. The Temporal Voice Areas (TVA's) have recently been localised in the bilateral temporal lobes, extending along the posterior (TVAp), mid (TVAm) and anterior (TVAa) STS and STG bilaterally, along with the involvement of extra-temporal areas (bilateral inferior prefrontal cortex and amygdalae) (Ahrens et al., 2014; Belin and Grosbras, 2010; Pernet et al., 2015a; Pernet et al., 2007). The voice localisation task showed bilateral activations over the temporal voice areas, with further activation of the left anterior STS/STG (Figure 5-7), similar to control data (Pernet et al., 2015a).

Phoneme perception and voice gender perception were investigated (Jones et al., 2015) using 2 AFC identification tasks: voice gender (male vs. female) and phoneme (/pa/ vs. /ta/). Using the classification criterion described in Chapter 4 (Jones et al., 2015), CF was classified as unimpaired on the phoneme perception task, and impaired on the voice perception task – the same performance as the majority of patients in the right fronto-temporal damaged patient group [8 out of 9 patients, χ^2=7, p=0, Φ=inf : Chapter 4 (Jones et al., 2015)] (Figure 5-7).

CF's tone discrimination threshold (Hz) was assessed using the interfering tone discrimination task (Jones et al., 2015) compared to the scores obtained by the control group assessed in Chapter 3 (Pernet et al., 2013) and the patient groups tested in Chapter 4 (Jones et al., 2015) and no significant difference was found (ANOVA) ($F_{(3,40)}=1.617$, p = 0.201).

Shortly following his stroke, CF was sent for a full hearing evaluation as it was imperative for this study, and for his SLT assessment and intervention, that any hearing loss was recorded, accounted for, and corrected. CF was found to have a bilateral, asymmetrical sensorineural hearing loss (Figure 5-6). His hearing loss ranges from a normal level in the lower frequencies, through to a moderate level of loss in the high frequencies (2KHz – 8KHz). CF was provided with bilateral hearing aids, which fully corrected his impairment. As CF's hearing loss was corrected for, and due to the fact that he was unimpaired on the tone discrimination task, it is not possible to attribute his impaired performance on the voice perception task to a hearing loss of any kind.

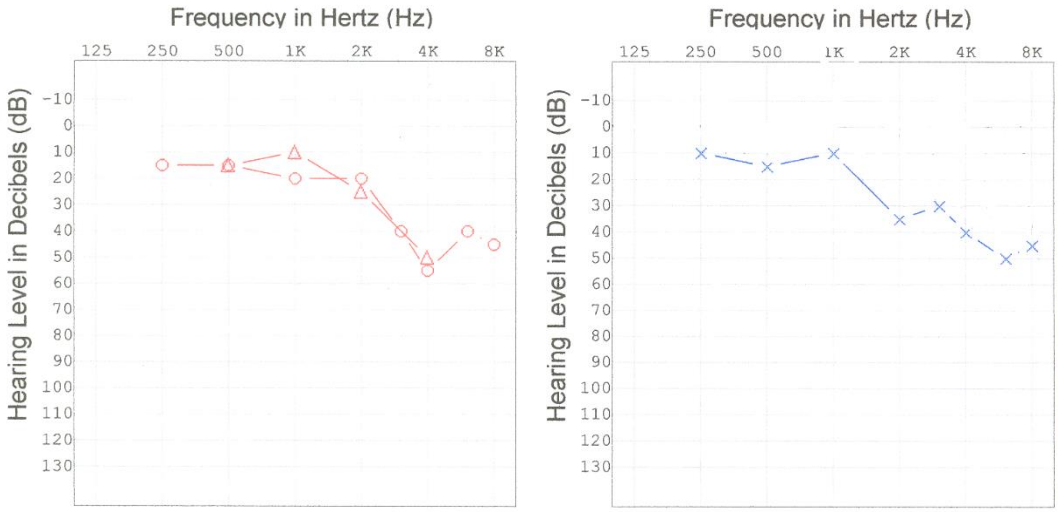

Figure 5-6: Audiogram carried out on CF showing bilateral, symmetrical sensorineural hearing loss. Left hand graph shows unmasked air conduction (circle) and unmasked bone conduction (triangle) assessment results carried out on the right ear; right hand graph shows unmasked air condition assessment results carried out on the left ear.

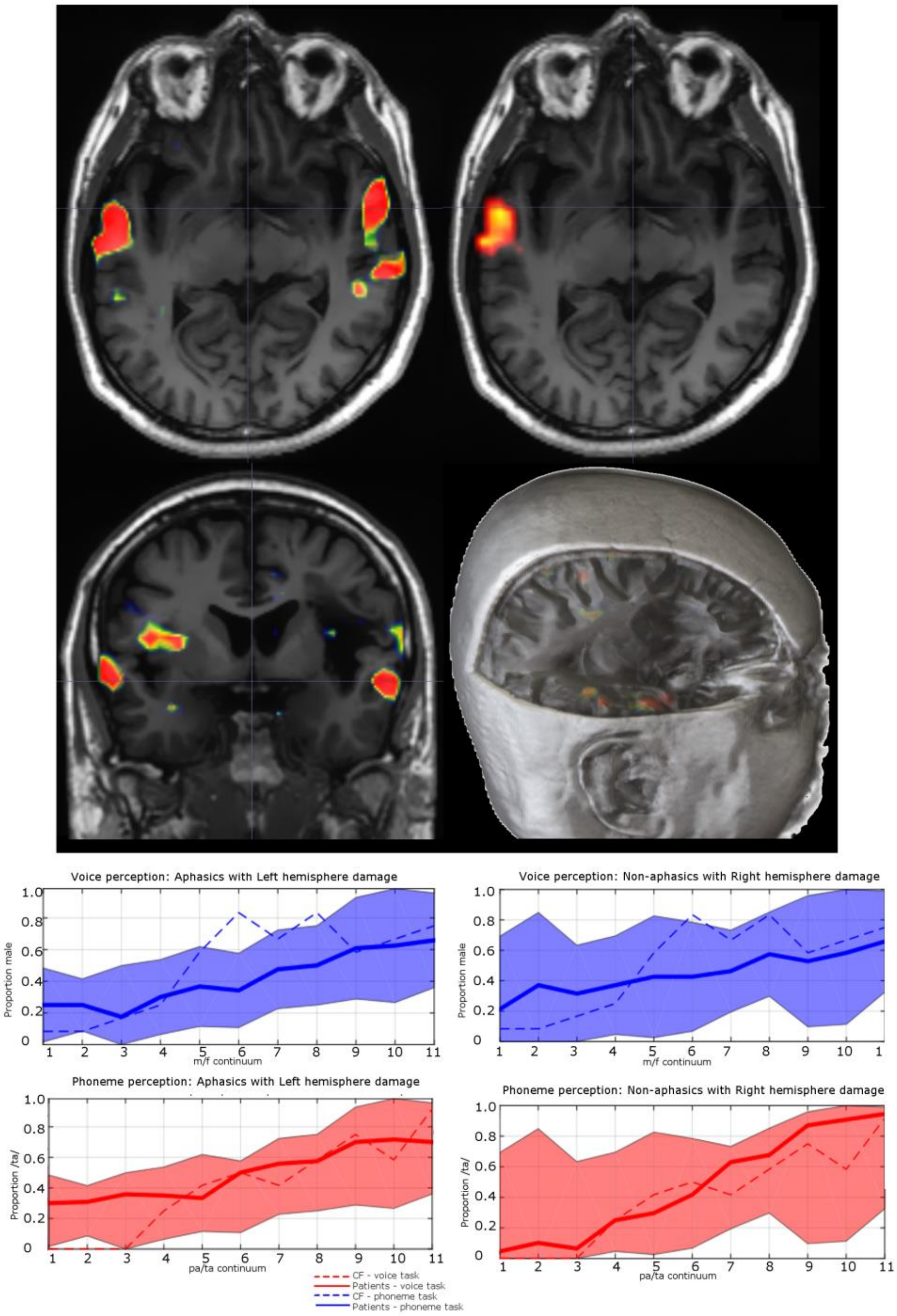

Figure 5-7: fMRI and behavioural results for voice and speech perception assessments

Imaging results: Single subject maps from voice perception fMRI task. Coronal (top) slices, axial (bottom left) slices and rendered (bottom right) images showing areas of activation and lesion. Left: alpha = 0.001; right: alpha = 0.005. Note left is on left, right on right (neurological convention).

Behavioural results: CF's performance on voice perception and phoneme perception tasks as compared to control patient groups (Left: Patients with left hemispheric damage and aphasia; Right: Patients with right hemispheric damage and no aphasia). Blue - voice perception task; Red – phoneme perception task; solid line – mean scores from a control group of patients. Shaded areas show confidence intervals for control groups (adjusted for multiple comparisons). Dotted lines show CF's performance at the chronic time point on each task.

Cognition

CF presented with variable memory abilities suggesting external influencing factors such as fatigue, mood and stress/anxiety. CF performances' were in the normal range for visuo-spatial skills and executive functioning. He did not show any signs of central apraxia even though there was damage to the MFG, showing that CF's praxis system was not damaged by the lesion. CF did not show any signs of visual field neglect either, possibly because the lesion did not damage his parietal lobe, remaining more anterior. Acalculia screening showed some impairment when carrying out simple additions and deletions (Appendix 12). These difficulties corresponded to numerical length, suggesting impairment in CF's output buffer rather than numeracy itself. He performed poorly on a specialist knowledge retention task, prompting further, detailed, analysis of the results of this informal assessment to determine whether this was a similar presentation to those with semantic dementia and associated aphasia (Graham et al., 1999a; Graham et al., 1999b; Hirono et al., 2000; Omar et al., 2010; Robinson et al., 1999) or due to other associated impairments. On examination it was apparent that his difficulties were in word-finding and formulation of sentences, and not due to the actual retention of expert semantic knowledge (Figure 5-5, Figure 5-8). (Appendix 13 shows cognitive profiles in full).

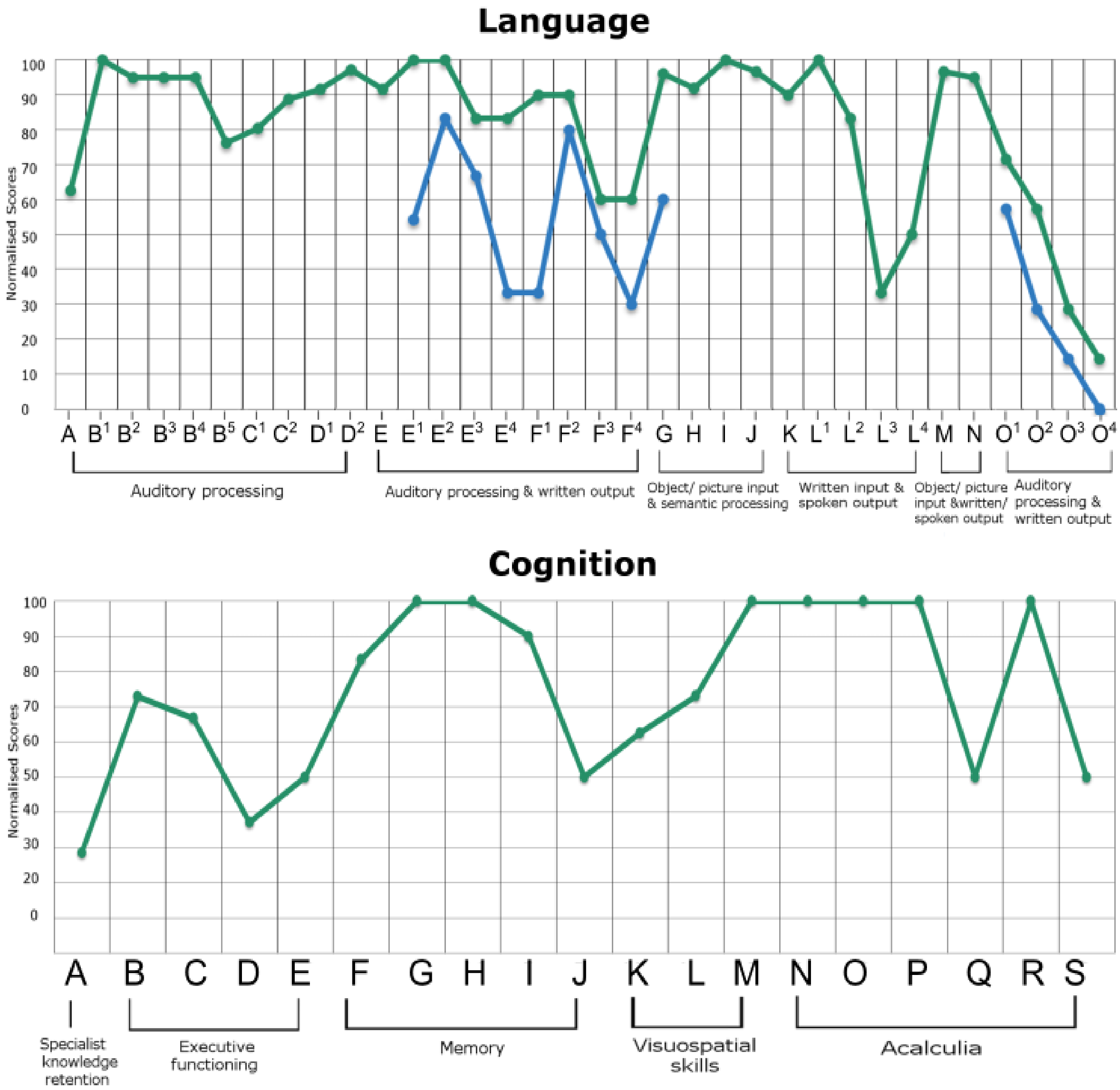

Figure 5-8: CF's language and cognitive profiles (language scores).

Blue line: acute phase of testing; green lines: chronic phase of testing. For language (top), normalized scores are presented for phonological discrimination (A), PALPA 5 (high/low imageability, high/low frequency: B1, B2, B3, B4; nonwords: B5), PALPA 1 (same/different judgements: C1, C2), PALPA 2 (same/different judgements: D1, D2), PALPA 39 (total: E; 3,4,5,6 letters: E1, E2, E3, E4), PALPA 40 (high/low imageability, high/low frequency: F1, F2, F3, F4), PPT (G), KDT (H), TTT (I), SPMT (J), PALPA 8 (K), PALPA 36 (3, 4, 5, 6 letters: L1, L2, L3, L4), PALPA 54 (written output: M; spoken output: N), PALPA 45 (3, 4, 5, 6 letters: O1, O2, O3, O4). For the other cognitive domains (bottom), normalized scores are presented for expertise retention (A), BADS (modified 6 elements: B; temporal judgement: C; key search: D); WAIS IV matrix reasoning (E); WMS II Faces I and II (visual immediate memory: F; visual delayed memory: G); BMIBP (total: H; speed of information processing: I), Rey-Osterrieth Complex figure (delayed recall: J; copying: K); VOSP (incomplete letters: L; dot counting:

M); numeracy (counting forwards: N; counting backwards: O; size decisions: P; simple addition: Q; simple multiplications: R; simple deletions: S).

5.2.4. Discussion

Crossed aphasia

Using Mariën et al's (2004) algorithm, it is concluded that CF is a reliable case of vascular crossed aphasia: he is a right-handed patient with no left-handedness in his family, showing morphological integrity of the left-hemisphere and no brain damage or seizures in childhood. CF shows clear-cut evidence of language disorder (anomic aphasia) with word-finding difficulties related to auditory phonological analysis. Functional MRI analyses show right hemispheric lateralization of language functions (reading, repetition, generation), concurring with results from the mini-literature review on cognitive co-morbidities in crossed aphasia. Central apraxia, visual field neglect and acalculia were significantly associated with CA, which can only occur if they are right lateralized (as in most individuals) along with language, including reading and writing systems, which explain the association with agraphia.

Voice and speech perception

Although CF was unimpaired in phoneme perception, as expected, he showed an impairment in voice perception – the same pattern as in patients with right fronto-temporal damage and no aphasia (Jones et al., 2015). Even though functional MRI analyses in CF show bilateral activations over the temporal voice areas (as seen in controls), the frontal damage led to deficits, which confirm the aformentioned hypothesis of a dissociation between voice and speech, even here in CA. The absence of phonological deficit, at the chronic phase, relates to CF's specific lexical difficulty (word-finding), which can co-occur with intact phoneme perception i.e. in bottom-up processing within the cohort model (Norris, 1994; Norris et al., 2000; Norris et al., 2001).

Dysgraphia and dyslexia

CF presented also strong dysgraphia which is believed to be a pre-existing, developmental disorder, not formally diagnosed due to lack of knowledge and diagnosis available at this time (Swanson et al., 2013). Developmental dysgraphia can be defined as "a specific learning disorder […] an impairment in written expression" (American Psychiatric Association, 2013) that causes problems in handwriting only, spelling only, or both handwriting and spelling (Berninger et al., 2001). CF reported problems from a 'young age' with spelling – i.e. he did not have any other developmental or medical conditions, and he presents with a breakdown in his graphemic output buffer (Figure 5-5), leading to difficulty spelling non-words and effects of word length, imageability and frequency - impairments involving 'writing specific' processes (Planton et al., 2013). CF's dysgraphia can further be classified as 'dyslexic aphasic dysgraphia' i.e. a language disorder mainly characterised by a writing impairments consisting of "mis-spellings with reversals, omissions, inversions and substitutions of non words and paragraphic errors" (Gubbay and De Klerk, 1995) consistent with an impairment in phoneme/grapheme correspondence. Impaired non-lexical reading was also present, also suggesting the presence of a mild phonological dyslexia, which concurs with the idea of a pre-morbid impairment in phoneme/grapheme correspondence.

Conclusions

The possible causes underlying crossed aphasia have been investigated for a number of years but, as only four of the 30 papers from this literature review (Alexander and Annett, 1996; Bakar et al., 1996; Cappa et al., 1993; Osmon et al., 1998) look into this, it is an area that requires further attention.

Putting back CF's deficits in the general context of CA, it can be postulated that CF had left hemisphere defects causing the dysgraphia and dyslexia and causing a right hemispheric language shift. As recent evidence suggests that lateralisation shift occurs not only with large lesions, but also small focal lesions or dysfunction of neural networks (Guerreiro et al., 1995; Kurthen et al., 1992; Lazar et al., 2000; Maesto et al., 2004; Staudt et al., 2001), it can be further concluded that CA can be caused by a congenital dysfunction within the left reading/writing systems, and not just the left auditory/spoken system (Bakar et al., 1996; Bhatnagar et al., 2006; Cappa et al., 1993). In addition, since (i) only 1 out of the 57 cases identified in the review could be conclusively attributed to a genetic basis (Cohen et al., 1993), (ii) bi-hemispheric representation was suggested in 10 of the remaining 56 cases (Bakar et al., 1996; Cappa et al., 1993; Giovagnoli, 1993; Habib et al., 1983; Ishizaki et al., 2012; Paghera et al., 2003; Paparounas et al., 2002), and (iii) when tested, dysgraphia co-occurred in >60% of CA cases, it is conceivable that developmental disorders cause a total or partial right lateralisation shift in language functioning, at least in some cases.

Advances in imaging techniques have allowed the investigation of the neurological basis of dyslexia (Cao et al., 2006; Habib, 2000; Shaywitz et al., 2002; Shaywitz et al., 1998; Shaywitz et al., 2003; Shaywitz et al., 2006). It has been shown that deficits in the left inferior frontal gyrus, left inferior parietal lobule (i.e. supramarginal gyrus & angular gyrus: Singh-Curry and Husain, 2009), and mid-ventral temporal cortex (i.e. fusiform gyrus, parahippocampal gyrus, lingual gyri & inferior temporal gyri: Haxby et al., 2001) are associated with developmental dyslexia in dextrals. Shaywitz (2002; 2003) has further suggested that such left hemisphere disruptions are compensated for by recruitment of the right hemisphere, supporting the theory that developmental disorders can be an underlying cause of crossed aphasia.

Also of importance, is the fact that CF shows a dissociation between voice and phoneme perception. Even though CF's voice perception impairment was not causing him problems at speech level, it may hinder other aspects of his communication such as his ability to gather identity (differentiate gender, identify individuals etc.) and affective (emotional and motivational states) information (Belin et al., 2004).

Finally, Berthier et al. (2011) do report a case of crossed aphasia in a patient with developmental dyslexia and dysgraphia, however this patient was ambidextrous and therefore cannot be classified as a confirmed case of crossed aphasia (Marien et al., 2004). Therefore the case of CF, presented here - a confirmed crossed aphasia with developmental dyslexia/dysgraphia - is the only known reported case.

6. Chapter 6 – Intact Single Word Auditory Comprehension

This study investigates the neural processes underpinning auditory comprehension in the healthy population using a word/picture verification task (Breese and Hillis, 2004).

6.1. Study 4 – Single word level auditory comprehension

6.1.1. Introduction

Oral presentations:

Jones, A. (2013). *Auditory Comprehension studies – ethical considerations.* Presented at University of Glasgow - School of Medicine: Medical Tutorial Series; 2013 January 29; Glasgow, UK.

Jones, A. (2012). *Auditory Comprehension: intact and impaired.* Presented at Chest Heart and Stroke Study Day: Technology; 2012 November 21; Edinburgh, UK.

Contributions by Anna Jones: study design, ethical submission, patient recruitment, data collection and analysis, results interpretation, writing up of the study, writing and giving oral presentations.

Impaired auditory comprehension, 'an inability to understand linguistic utterances which cannot be attributed to deficient sensory input, generalised cognitive defective, or defective attention' (Boller et al., 1977), is frequently assessed using multiple choice tasks, i.e. Boston Diagnostic Aphasia Examination (BDAE) (Goodglass et al., 2001) or sections of the Psycholinguistic Assessment of Language Processing in Aphasia (PALPA) (Kay et al., 1992). However, word/picture verification tasks (Breese and Hillis, 2004; Hillis et al., 2002) have a higher sensitivity for identifying impairments in auditory word comprehension than multiple choice tasks (Breese and Hillis, 2004). In addition, error patterns in word/picture verification tasks can be analysed, giving information on particular stages of disrupted lexical processing and thus providing useful information for clinical and research purposes.

During the word/picture verification task participants are presented each picture 4 different times with different auditory words: once with the correct name (‘target’), once with a phonologically related word (‘phonological foil’ e.g. cat/hat), once with a semantically related word (‘semantic foil’ e.g. cat/dog) and once with an unrelated word (‘unrelated foil’ e.g. cat/ house). In this type of task individuals must accept the target correctly and reject all the foils in order for the item to be scored as correct. Even when the same number of foils are presented in a multiple choice task the chance of correctly guessing is 25% whereas in the word/picture verification task the chance of guessing correctly is reduced to 6.25% ($50\%^4$).

The word/picture verification task is based on the interactive distributed model of semantic processing where semantic memories are stored as distributed representations of semantic features. This semantic processing model, ‘Hopfield Networks’ (Hopfield, 1982), is a modification of the connectionist network models of spoken word recognition and auditory comprehension discussed in Chapter 1, for example the Logogen model, the TRACE model and the Adaptive Resonance Theory (Grossberg, 1980; McClelland and Elman, 1986; Morton and Patterson, 1980). In this model individual semantic features are represented by different nodes meaning that no two items will activate an identical set of features. Therefore, when an individual is accessing an underspecified semantic representation both the target and semantic foil are usually selected, causing an error only on the foil. In multiple choice tasks with 3 foils the forced guess chance rate is 25%, whereas on word/picture verification tasks (again with 3 foils) forced guessing is eliminated and usually the participant will accept both the semantic foil and the target, causing an error only on the semantic foil. Therefore, it is possible to use error patterns on the word-picture verification task to investigate impairment levels on the cognitive neuropsychological model of single word processing detailed in Chapter 2 (Figure 2-8). This model is an adapted and extended version of Ellis et al.’s (Ellis and Young, 2004) original Transcoding Model, which build upon earlier connectionist models such as the Logogen model (Morton and Patterson, 1980).

Although a number of studies have been carried out using picture/word verification tasks as measures of auditory comprehension, only one study has actually looked at the task in detail (Breese and Hillis, 2004). Furthermore, although the neuroanatomy of auditory comprehension has been investigated in detail and models produced (i.e. the triple-stream auditory comprehension model - (Specht, 2014) – See Chapter 1, Figure 1-8), limited studies have looked into the neural regions and networks involved in auditory single word comprehension (left superior temporal gyrus - (Hillis et al., 2002)) at different levels i.e. phonological, semantic.

6.1.2. Method

Ethics Statement

This study was approved by the NHS Scotland A Research Ethics Committee (*REC reference number: 11/AL/0069*) and NHS Lothian Research and Development (*R&D project number: 2011/W/ST/01)* and Edinburgh Clinical Research Facility (*CRF approval number: E11928 CRF*). All participants were provided with detailed information regarding the nature of the study and their potential role in it and full informed consent was obtained from participants (see Appendix 15 and 16).

Participants

16 healthy subjects (8 male, 8 female; mean age 31 years, mean number of years of education 20) participated in this study. All participants were right-handed, as assessed using the Edinburgh Handedness Inventory (Oldfield, 1971). Participants were excluded if they had any uncorrected visual or hearing impairment, any known neurological disease or disorders, or English was not their first language.

Procedure

All participants carried out a single word auditory comprehension task, based on Breese & Hillis' (2004) spoken word/picture verification task. This task aimed to map auditory comprehension areas of the brain, and specifically areas engaged in phonological and semantic comprehension of spoken nouns. During the task, a series of 30 pictures were each presented 4 times, along with a spoken name. Each time the spoken name associated to the image was different: once it matched the picture (correct target, e.g. 'cat'), once it was phonologically related (phonological foil, e.g. 'hat'), once it was semantically related (semantic foil, e.g. 'dog') and once it was unrelated (unrelated foil, e.g. 'spoon'). The task was designed to minimise any confounding effects of spoken output, written input and written output, as these modalities were not required to complete the task. Importantly, memory load during the task was minimal as the pictorial and auditory representations of the word were presented at exactly the same time. As all the words used were real nouns this allowed assessment of the phonological input lexicon and the semantic system, thereby enabling identification of any single word auditory comprehension impairments.

Note that throughout this study the correct target is also referred to as the ‘congruent’ condition (i.e. the sound and picture stimuli accurately match) whereas the phonological, semantic and unrelated foils are also referred to as ‘incongruent’ (i.e. the sound and picture stimuli do not accurately match and are thus different in terms of their phonology, semantics or are completely unrelated).

Stimuli

Stimuli chosen were all single nouns, deemed to be culturally appropriate and no stimuli were repeated during the task. All nouns were taken from the British National Corpus (BNC Consortium, 2007). Target nouns were chosen according to frequency, imageability and familiarity ratings (BNC Consortium, 2007). The phonological foils all differed from the target by a maximum of two phonemes. Semantic foils were related to the target by category and unrelated foils were nouns not related to the target, semantic foil or phonological foil in either a semantic or phonological level. All foils were also chosen according to their frequency, imageability and familiarity ratings (BNC Consortium, 2007).

Pictures were taken from the 'Bank of Standardized Stimuli’ (BOSS) (Brodeur et al., 2010) and were only included in the study if their name agreement (between noun and picture) was above 75%. All sound stimuli were each recorded by a single female voice in a sound studio at the Voice Neurocognition Laboratory (http://vnl.psy.gla.ac.uk/).

Each sound file was normalised based on the lowest energy measured among all the samples using a root mean squared (RMS) function, bringing the average amplitude to a target level (1KHz). Stimuli were presented in a pseudo-random order according to imaging parameters.

Imaging

Imaging was obtained using a GE Signa HDxt 1.5T scanner at the Brain Research Imaging Centre (http://www.bric.ed.ac.uk/), University of Edinburgh, UK. Structural imaging consisted of a T1-weighted volume (3D IRP - 160 slices, 1.3 mm thick coronal slices, 1.3 x 1.3 mm in-plane resolution with a 256 mm FOV). fMRI was performed to map single word auditory comprehension areas (EPI – 30 slices, 4mm thick axial slices, 4 x 4 mm in-plane resolution with a 256 mm FOV, TR = 2500 msec, TE = 50 msec).

The auditory comprehension task was implemented using a stochastic event-related paradigm with a jittered inter-stimulus-interval of 4, 6 or 8 seconds (Figure 6-1 – section C), and lasting 12 minutes in total. A total of 120 words/pictures were delivered (pseudo-randomly) during the task, and each stimulus was shown for 2 seconds. Participants were instructed to remain as still as possible, moving only their left index finger or left thumb to respond, throughout the scanning procedures. Stimuli presentation was delivered using the Psychophysics Toolbox Version 3 (PTB-3) (2014; Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) under Matlab (2014) and synchronization and stimulus presentation achieved using Nordic Neurolab® equipment.

Participants were instructed as follows:

“You will see pictures and hear sounds. Some will match, some will be different. Please press the button with your THUMB if they MATCH. Please press the button with your FINGER if they are DIFFERENT. You are now ready to start the practice…”

Following a practice trial the responses were collated and only when the participant reached a rate of 90% correct were they instructed that the main task would commence. If their score was >90% they commenced the trial again.

6.1.3. Data Analysis

Behavioural analysis

Overall performances for each subject were calculated out of 30 (i.e. to score 1 point the participant had to correctly answer that the target matched and that all the foils did not match). A robust one sample t-test was performed on the group results to test if the mean score is different from the maximum score (30).

Scores (%) and reaction times (RTs – sec) for each condition (correct/ phonological/ semantic/ unrelated) were calculated, per participant. Differences between the conditions (for both RTs and scores) were investigated using robust repeated measure ANOVAs and any incongruency effects (i.e. matched conditions vs. unmatched conditions) and differences between the incongruency effects (i.e. differences between the unmatched conditions) were investigated by performing multiple comparisons between data pairs (percentile bootstrap on differences).

fMRI imaging analysis

Preprocessing

The images were all analysed using the Statistical Parametric Mapping software (SPM12, Wellcome Trust Centre for Neuroimaging, London, UK, http://www.fil.ion.ucl.ac.uk/spm/). First all scans were manually reoriented to the intercommissural (ACPC) line. Second, data were pre-processed using default parameters: slice timing correction, realignment to correct for any head movement, co-registeration of the T1-weighted image to the mean EPI images, segmentation and spatial normalisation to MNI space.

Statistical analysis

Statistical analyses were performed using the general linear model (GLM), implemented in SPM 12 (http://www.fil.ion.ucl.ac.uk/spm/), whereby four conditions (correct, phonological, semantic and unrelated) were modelled as four regressors (plus error responses and motion regressors) convolved with the HRF and its derivatives to investigate activations whilst accounting for the variability in the haemodynamic response across subjects and brain regions. A parametric modulation was also applied to allow for modulation of the activations with reaction times (RTs). Statistical parametric maps were generated for the different experimental conditions, 'boosted' by combining basis functions (Calhoun et al., 2004; Pernet, 2014), and smoothed with a 6mm FWHM isotropic Gaussian kernel. Group analysis was then undertaken looking first at the overall task effect, and activations for each condition (one-sample t-tests on each condition boosted beta parameters as well as on the boosted contrast for all conditions and all parametric conditions). Finally, a 2-way repeated measures ANOVA with the 4 conditions (correct, phonological, semantic and unrelated) was performed on the boosted parameter estimates to assess the likelihood of differences between the target and the foils (i.e. looking at congruency and incongruency effects).

6.1.4. Results

fMRI behavioural results

Performance on the auditory comprehension task was uniformly good, with participants attaining an average overall performance score of 26.19 +/- 2.01 (max. 30). Accuracy rates per condition ranged from an average of 99.33% +/- 1.45 (semantic) to 92.75% +/- 6.02 (phonological) and reaction times per condition ranged from 0.96 sec +/- 0.30 (correct) to 1.21sec +/- 0.19 (phonological) (Table 6-1 & Figure 6-1 – section B).

	Overall performance (max. 30)	Accuracy rates (%)				Reaction times (sec)			
		C	P	S	U	C	P	S	U
mean	26.19	98.62	92.75	99.33	93.30	0.96	1.21	1.13	1.07
S.D.	2.01	2.35	6.02	1.45	2.91	0.30	0.19	0.21	0.20

Table 6-1: Mean (with standard deviations) overall performance and accuracy rates (%)/ reaction times (sec) per condition.
C: correct target; P: phonological foil; S: semantic foil; U: unrelated foil.

There was a significant difference of accuracy rates between the correct condition and the foil conditions: F(3,13)=9.479, p=0.024, with a significant incongruency effect for the phonological conditions (-5.87%, p=0.004). In terms of reaction times there was also a significant difference between the target and foil conditions: F(3,13)=30.649, p=0 with a significant incongruency effect for the phonological conditions (+250ms, p=0) and the semantic conditions (+170ms, p=0) (Table 6-2 & Figure 6-1 – section D, left side).

Significant differences were also seen between the incongruent effects in terms of accuracy rates [F(2,14)=9.430, p=0.008] and reaction times [F(2,14)=14.823, p=0]. The phonological incongruency effects were stronger than the unrelated incongruency effects, and the phonological incongruency effects was stronger than the semantic incongruency effect but for accuracy rates only (Table 6-2 & Figure 6-1 – section D, right side).

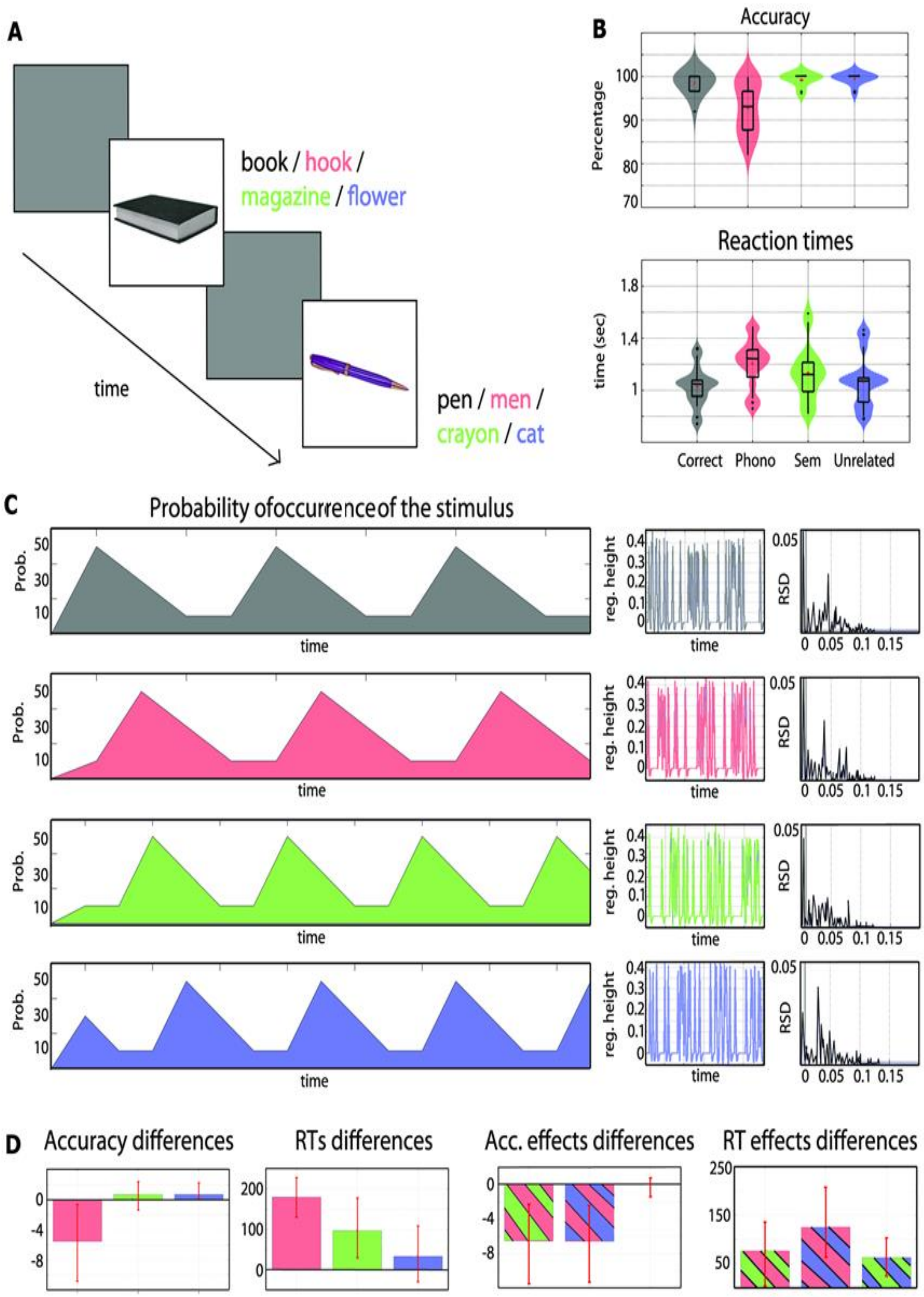

Figure 6-1: Behavioural results and design for fMRI auditory comprehension task.

Colours refer to the 4 conditions: grey- correct target; pink- phonological foil; green- semantic foil; blue- unrelated foil.

A: Event-related semi-random presentation of stimuli. Each of the 30 target nouns were presented pictorially 4 times throughout the task, once with each of the 4 conditions. In total 120 stimuli were presented for 2 seconds each, with 4, 6 or 8 seconds intervals in between, over a total time of 12 minutes.

B: Violin plots (boxplot showing IQ range with a rotated kernel density plot on each side to signify the probability density of the data at the different values (Hintze and Nelson, 1998)) showing behavioural results per condition for accuracy rates (%) and reaction times (sec), with the median data point (red dot) overlaid.

C: Left shows the probability of the occurrence of stimulus from each condition in the fMRI presentation. Right column shows the height of the regressors (contributors) across scanning time, with the associated relative spectral density (RSD) across frequency (Hz) for each condition. This shows a good spread of signal across a range of frequencies as well as a similar power efficiency throughout the conditions.

D: Pair-wise comparison graphs showing the differences between pairs of conditions with confidence intervals shown (red – 98.31%) for (left-right):

- accuracy rates for correct (congruent) vs. incongruent conditions (phonological – semantic – unrelated);
- reaction times for correct (congruent) vs. incongruent conditions (phonological – semantic – unrelated);
- accuracy rates between incongruent conditions (phonological vs. semantic – phonological vs. unrelated -semantic vs. unrelated);
- reaction times between incongruent conditions (phonological vs. semantic – phonological vs. unrelated -semantic vs. unrelated).

	Accuracy rate					
Condition pairs	**C/P**	**C/S**	**C/U**	**P/S**	**P/U**	**S/U**
Difference	*-5.538*	0.678	0.678	*-6.521*	*-6.561*	0
CI	*[-11.16 -0.84]*	[-1.07 2.24]	[-0.01 2.18]	*[-11.45 -1.75]*	*[-11.10 -1.75]*	[-1.46 0.74]
p-value	*0.004*	0.638	0.450	*0*	*0*	0.226
	Reaction times					
Condition pairs	**C/P**	**C/S**	**C/U**	**P/S**	**P/U**	**S/U**
Difference	*179.451*	*96.415*	32.637	*76.46*	*124.91*	*62.45*
CI	*[128.71 226.00]*	*[32.22 179.79]*	[-26.83 106.49]	*[9.80 133.83]*	*[67.16 209.12]*	*[23.34 103.88]*
p-value	*0*	*0*	0.214	*0.006*	*0*	*0*

Table 6-2: Percentile bootstrapped pair-wise differences on accuracy rates and reaction times between matched conditions and unmatched conditions.

C: correct target condition; P: phonological foil condition; S: semantic foil condition; U: unrelated foil condition. Significant differences shown in italics.

fMRI imaging results

All group activations were considered significant after multiple comparisons correction using a cluster size false discovery rate (cluster forming threshold p=0.001, cluster FDR q<0.05).

Group activations - all conditions:

The overall task effect (one sample t-tests using the constant (Figure 6-2; right) and parametric parameter estimates (Figure 6-2; top right) showed activations in the left inferior occipital cortex, including the left fusiform gyrus, bilateral supplementary motor cortex and the bilateral inferior frontal gyri. Only constant effects were observed in the bilateral superior temporal gyrus (STG) and the right fusiform gyrus (FG). Only parametric effects were observed within the bilateral insula, the left supramarginal gyrus (SMG/ Wernicke's area), the left superior parietal lobule, and the right middle frontal gyrus. Deactivations were significant with the constant parameter estimates pooling all conditions within the left angular gyrus, the right hippocampus and the anterior/posterior cingulate.

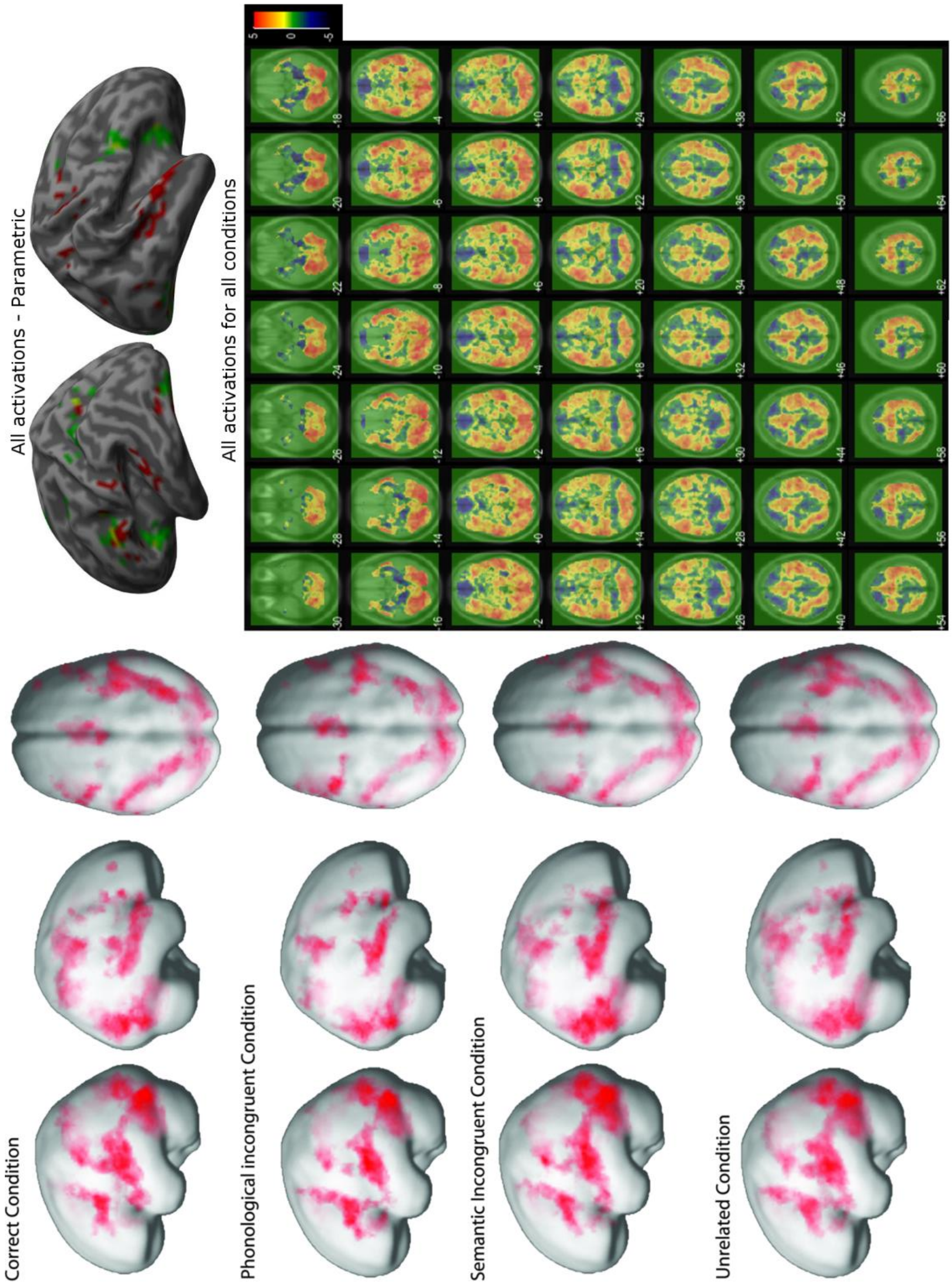

Figure 6-2: Group results from the single-word auditory comprehension fMRI task.

Left: 3D rendering of activations (hrf boosted) from the group analysis for each condition, projected onto the gray-matter MNI space template from SPM12. FDR corrected: correct (threshold 0.05, k=94); phonological (threshold 0.05, k=55); semantic (threshold 0.05, k=52); unrelated (threshold 0.05, k=50). *Top right:* 3D rendering of activations (hrf boosted) from the group analysis for all conditions together (red) and for the parametric analysis (green), projected onto the canonical cortex template from SPM12. FDR corrected (threshold 0.001, k=100).

Bottom right: Raw activations map showing unthresholded activations (hrf boosted) from the group analysis for all conditions together, overlaid onto axial slices (-30:2:66) of the canonical average T1 template from SPM12.

	Anatomical location	MNI co-ordinates			$P_{FWE\text{-}corr}$	$P_{FDR\text{-}corr}$	Z-score
		x	y	z			
C > P	Left Precentral gyrus	-12	-24	48	0.024	0.038	4.59
	Medial cingulate	4	12	36	0.021	0.038	4.04
C < P	Left Pars triangularis	-52	30	8	0.000	0.000	5.71
C > S	Left Precuneus	-8	-74	28	0.015	0.015	5.22
	Left Angular gyrus	-42	-54	44	0.006	0.013	4.86
C < S	*no activations of significance*						
C > U	Left Angular gyrus	-44	-56	44	0.043	0.077	4.50
C < U	*no activations of significance*						

Table 6-3: Overview of all large clusters of activity showing the main effects of the ANOVA, along with the co-ordinates of local maxima, anatomical locations and z-scores.
Corresponding anatomical locations are shown, as determined by the Neuromorphometrics atlas (http://www.neuromorphometrics.com/), the SPM anatomy toolbox (Eickhoff et al., 2005) and/or visual inspection. Local maxima were thresholded at 0.001 uncorrected and < 0.05 whole brain FWE and FDR corrected. C: correct condition; P: phonological condition; S: semantic condition; U: unrelated condition.

Group activations – per condition:

All conditions showed activations of the bilateral STG, bilateral fusiform gyrus, bilateral precentral gyrus, bilateral supplementary cortex, bilateral occipital cortex and the right angular gyrus. The left angular gyrus was also significantly activated in the phonological condition. Further significant activations were seen in the anterior insula – bilaterally for the phonological and unrelated conditions and mainly in the left hemisphere for the semantic condition. Significant activations were also seen in the SMG (Wernicke's area) in both the correct (left hemisphere) and unrelated conditions (right hemisphere). Furthermore, the right lingual gyrus showed significant activations in both the phonological and semantic conditions. Finally, the unrelated condition also showed activations in the right inferior frontal gyrus (IFG), not seen in the other conditions (Figure 6-2; left).

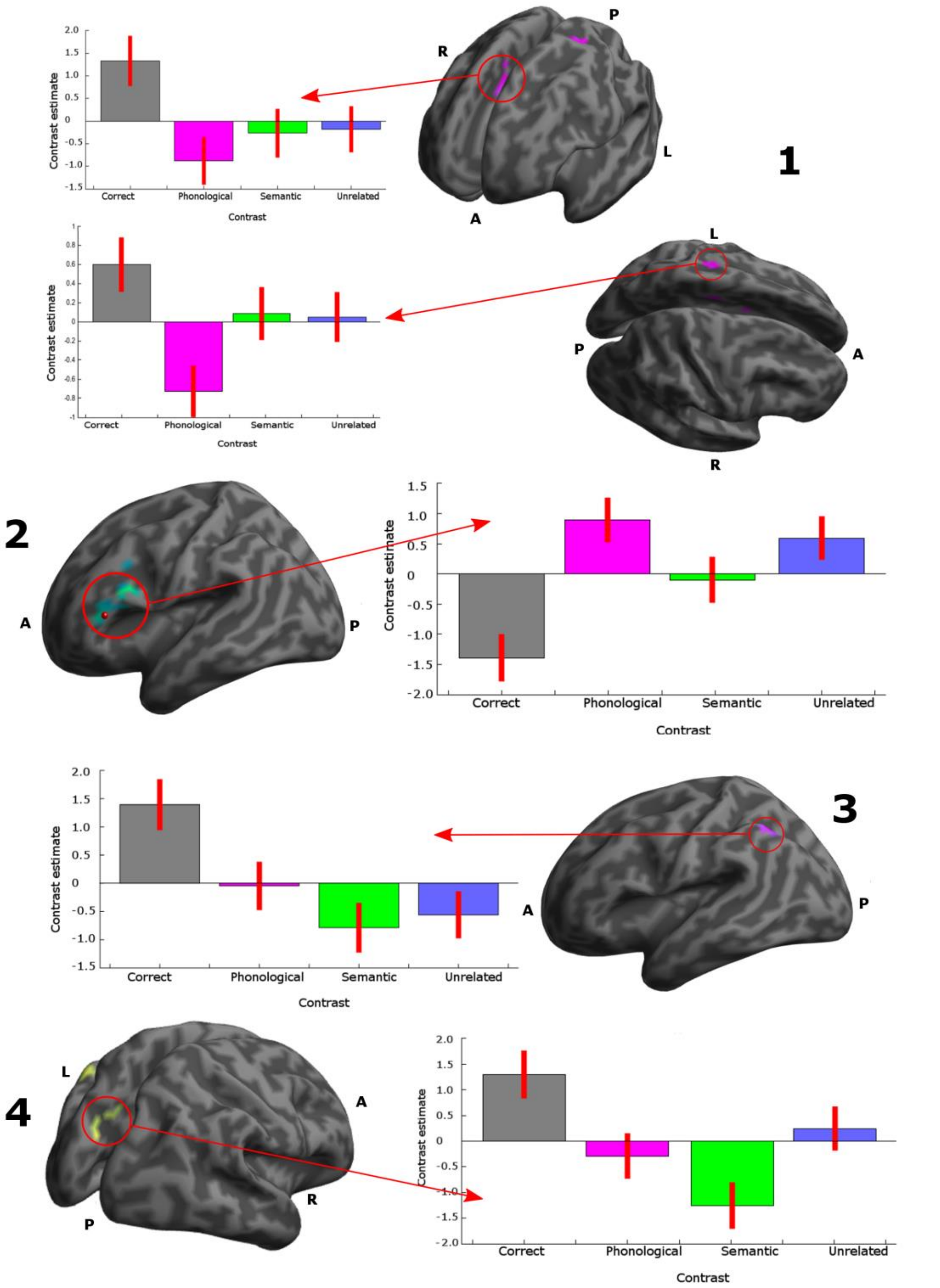

Figure 6-3: 3D rendering of activations (hrf boosted) and associated contrasts estimates from healthy participants (n=16) for the single word auditory comprehension task for congruent vs. incongruent effects.

Activations are projected onto canonical cortex from SPM12. FDR corrected (threshold 0.001, k=90 [min]). Contrast estimates and 95% confidence intervals are shown for each condition at the voxel of greatest activation (see

Table 6-3). A: anterior; P: posterior; L: left hemisphere; R: right hemisphere.

1: correct condition relative to the phonological condition (C>P) showed greatest activations in the medial cingulate (top) and the left precentral gyrus (bottom). *2:* phonological condition relative to the correct condition (C<P) showed greatest activation in Broca's area (left pars triangularis). The associated contrast estimate shows that Broca's area was also activated in the unrelated relative to the correct condition (C<U). *3:* both the correct condition relative to the semantic condition and the correct condition relative to the unrelated condition (C>S and C>U) showed greatest activation in the left angular gyrus. *4:* correct condition relative to the semantic condition (C<S) also showed activation in the left precuneus.

The repeated measures ANOVA revealed that the greatest overall effects (i.e. differences between the congruent and incongruent conditions) were in left angular gyrus and left inferior frontal gyrus (both the pars triangularis and pars opercularis). In both the medial cingulate (Table 6-3 and Figure 6-3, 1a) and the left precentral gyrus (Table 6-3 and Figure 6-3, 1b), significant differences were observed between conditions, showing an increased BOLD response in the correct conditions vs. no activation in the semantic and unrelated conditions (no significant differences) vs. deactivations in the phonological conditions (significant difference). In Broca's area (left triangular part of the inferior frontal gyrus – pars triangularis) (Table 6-3 and Figure 6-3, 2) significant differences were observed between conditions, showing an increased BOLD response in the phonological conditions vs. a reduced but still activated response in the unrelated condition (no significant difference in P>U for contrast estimates) vs. no activation in the semantic condition (no significant difference) vs. deactivations in the correct condition (significant difference). In the left angular gyrus (Table 6-3 and Figure 6-3, 3) significant differences were observed between conditions with an increased BOLD response in the correct condition vs. no activation in the phonological condition (no significant difference) vs. deactivations in the semantic and unrelated conditions (significant differences). Once again, in the left precuneus (Table 6-3 and Figure 6-3,4) significant differences were also observed between conditions, with an increased BOLD response in the correct conditions vs. no activations in the semantic and unrelated conditions (no significant difference) vs. deactivations in the phonological condition (significant difference).

6.1.5. Discussion

As all participants within this study had no uncorrected visual or hearing impairment, no known neurological disease or disorders, and all had English as their first language, it was hypothesised that there would be no distinction in accuracy rates between the conditions on the single word/picture verification task. However, results showed that subjects were more accurate on all conditions compared to the phonological foil. Reaction times were also slower on the phonological foil compared to the other conditions, together suggesting that the phonologically incongruent condition was the most difficult for participants. In terms of reaction times the semantic foil was also significantly slower than the unrelated condition, even though there was no overall significant difference in accuracy rates.

Following our model of single word processing (Figure 6-4), when presented with stimuli during the word/picture verification task the participant sees the picture, carries out object/ picture recognition (i.e. access to the structural description of the picture/object), and accesses object concepts/semantic system (i.e. access to the meaning of the picture/object), and the phonological form is then retrieved (Hillis, 2007). At the same time, the participant hears the spoken word, carries out auditory phonological analysis (using the phonological input buffer) then accesses the phonological input lexicon, and then the semantic system. While the model suggests that this occurs mainly in a serial manner, there is also some evidence that these systems interact and are activated in parallel, or in a cascade form (Rapp and Goldrick, 2000). At any stage of this process the participant can use the information s/he has gathered to make a decision as to whether the auditory and visual input are congruent or incongruent. Using the reaction time results from behavioural data it is possible to interpret the parametric/ non parametric imaging results and thereby inform our model of processing. As reaction times were slower on the phonological foil compared to the other conditions, it is suggested that when provided with the phonologically incongruent foil the phonological lexicon is accessed after the semantic system, and that the level of activation within the phonological lexicon is higher than that of the semantic system as the phonology is incongruent. When provided with the semantically incongruent foil, again both the semantic system/object concept system and the phonological lexicon are activated (in this order), however the levels of activation are reversed and the object concept/semantic system is activated more highly than the phonological system because the semantics are incongruent.

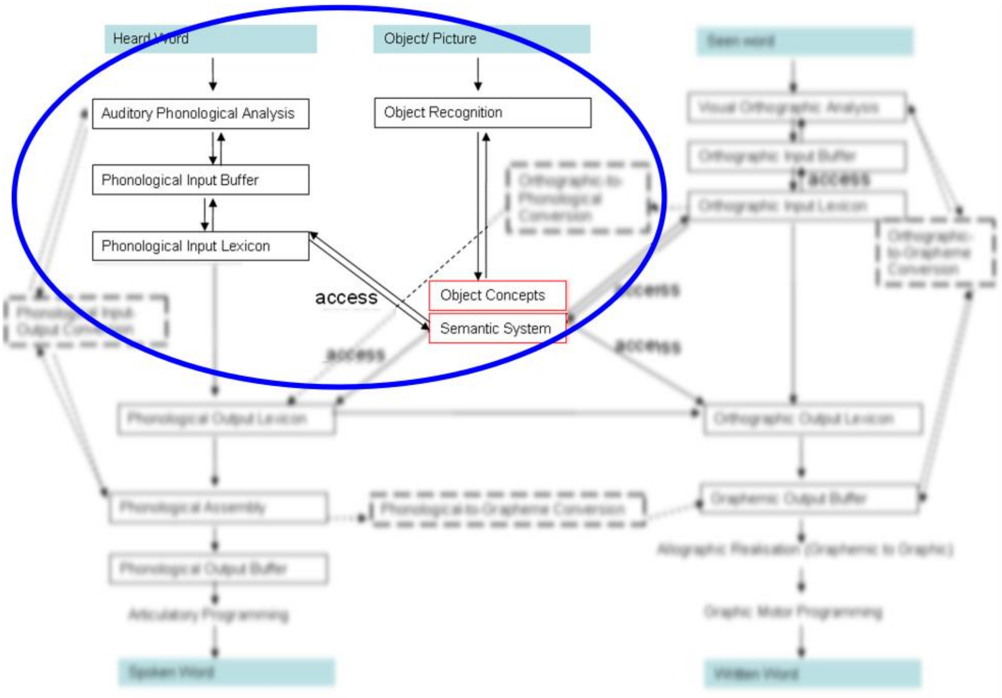

Figure 6-4: Picture/word verification task to assess single word comprehension according to the adapted and extended cognitive neuropsychological model of single word processing seen in Figure 2-8.

Phonological processing

During the phonological condition within this task, participants were required to make a decision when accessing both the object/picture recognition and the phonological input. From the fMRI results it is possible to see that the pars triangularis shows greater activation during the phonological condition than in the control condition. It is known that the pars triangularis (Brodmann's area 45) is located in the posterior portion of the left inferior frontal gyrus (IFG) and makes up part of Broca's area – commonly known for its key role in various speech and language functions and traditionally thought to play a selective role in semantic processing. There is also evidence supporting the role of the pars triangularis in phonological processing (Demonet et al., 1994; Fiez et al., 1995; Poldrack et al., 1999; Zatorre et al., 1996), supporting the anterior dorsal stream (articulatory network) of the 'triple-stream model of speech processing' ((Specht, 2014) – see Chapter 1, Figure 1-8). It is possible that during the phonological condition subjects were using sequencing and rehearsing strategies when processing the phonemic strings – an interpretation which is also consistent with motor theory of speech perception ((Demonet et al., 1992; Liberman and Mattingly, 1985) – see Chapter 1). Furthermore, there is emerging evidence showing the link between the pars triangularis and the phonological working memory loop (Rogalsky et al., 2008; Vigneau et al., 2006). This is a particularly compelling argument for the involvement of the pars triangularis in the phonological condition over the correct condition (which show deactivation) as the participant must make a decision at this level based on the closeness of the phonology to the target, requiring use of the phonological working memory, hence also the small activation in the unrelated condition and the overall modulation of the activity by the RT. Interestingly, while Broca's area shows trial to trial variations, thus reflecting stimulus-related information accumulation and decision making, such modulation was totally absent from temporal regions. This suggests that basic phonological analysis in itself doesn't affect the task, at least in healthy participants.

The question remains as to why the results showed significantly more activation in the left precentral gyrus and the medial cingulate for the correct relative to the phonological condition. The medial cingulate contains the cingulate motor areas (Vogt et al., 2003) - responsible for motor response selection and initiation (Hoffstaedter et al., 2012). Closer inspection of the data revealed that the activation on the precentral gyrus is actually anterior to the 'motor strip' (Hauk et al., 2004) – on the border between Brodmann's areas 3,2,1 (primary somatosensory cortex) and 4 (primary motor cortex). Interestingly, both the somatosensory system and the motor system have recently been implicated in the perceptual processing of speech, in terms of the motor theory of speech perception (Devlin and Watkins, 2007; Hickok and Poeppel, 2000; Hickok and Poeppel, 2007; Ito et al., 2009). Therefore, although these results link between the medial cingulate/ primary somatosensory cortex/ primary motor cortex with the motor theory of speech perception (Devlin and Watkins, 2007), it remains difficult to interpret why there is higher activation in the correct condition relative to the phonological condition in these particular areas.

Semantic processing

Results showed a significantly greater degree of activation within the left angular gyrus in the correct condition as compared to both the semantic and the unrelated conditions. Both the semantic and unrelated conditions are semantically incongruent and therefore the decision process for the individual occurs at the level of the semantic system (Figure 6-4), for both conditions. The left angular gyrus is known to be involved in semantics and in particular in lexico-semantic systems (Demonet et al., 1992). Anomic aphasia has long been attributed to lesions at the left temporo-parieto-occipital junction, including the angular gyrus. Once again, this corresponds to Spect's (2014) triple-stream model of speech processing whereby the ventral stream involves the left angular gyrus to carry out, amongst others, semantic processing. Importantly, the results support the fact that the angular gyrus is involved in the incongruency of semantics and not the degree of relatedness. Therefore, in the correct condition the stimuli are semantically identical causing an activation of the angular gyrus. This proposed functionality corresponds with the interactive distributed model of semantic processing on which the word/picture verification task is based (Hopfield, 1982), whereby semantic memories are stored as distributed representations of semantic features. Following the same reasoning, the absence of semantic similarity between auditory and visual stimuli should lead to deactivations, observed here for the semantic and unrelated foils. It should also be the case for the phonological foil but no activation or deactivation was observed. As for Broca's area, the angular gyrus activations (Figure 6-3, top right) vary according to reaction time responses, reflecting information accumulation and decision making.

Results also showed that the control condition activated the left precuneus significantly more than in the semantic condition, a result that does not show up in the control condition relative to the unrelated condition. Evidence suggests that the left precuneus is involved in a wide variety of different tasks that are inter-linked, including visuo-spatial imagery, self-processing (as part of the default network), as well as more posterior regions linked to episodic memory processing (Cavanna and Trimble, 2006). More recently the precuneus has been linked to the degree of the semantic relationship/ visual imagery (Kotz et al., 2002). Therefore, it is proposed that the more closely semantically related the stimuli are, then the higher the degree of visual imagery (a function of concreteness for semantically related words). Therefore, as this study used only concrete (not abstract) nouns, it is suggested that the semantically congruent condition (correct) activates the left precuneus more than the semantically related condition.

Using this task in the clinical population

This pilot study was conducted on a healthy population and therefore difficulties that may arise when using the same task on a clinical population must be addressed. Firstly, it is possible that a picture/word verification task may actually be detecting visual perceptual impairments instead of single word auditory comprehension impairments, therefore it is important to assess future participants with a battery of cognitive assessments to enable exclusion of such perceptual difficulties. Secondly, this task does not provide a strong assessments of semantics of actions (verbs) as only concrete nouns are used, and therefore a battery of detailed semantic tasks should be included when assessing a clinical population. Finally, poor results on this task may be due to a number of different impairments including difficulty accessing complete semantic representations, impaired single word recognition (access to phonological input lexicon), or impaired picture/ object representation. Thus it is important, especially with a clinical population, to carry out an extensive lexical battery to identify, and if necessary to rule out, such impairments.

Conclusions

This study shows: (i) that the pars triangularis is involved in phonological processing, most likely phonological working memory; and (ii) that the left angular gyrus is involved in semantic incongruency decisions, whereas the left precuneus is linked more specifically to semantic relatedness. Therefore, as these regions are often involved in aphasia, this task is useful in addressing specific phonological vs semantic impairments.

7. Chapter 7 – Impaired Single Word Auditory Comprehension

This study investigates impaired auditory comprehension at the level of the single word in post-stroke aphasia using a word/picture verification task (Breese and Hillis, 2004). A pilot study, exploring the use of computer-assisted therapy (React2©) in acute/ sub-acute post-stroke aphasic patients, is also presented.

7.1. Study 4 – Impaired single-word level auditory comprehension and an exploration of the use of computer-assisted therapy

7.1.1. Introduction

Oral presentations:

Jones, A. (2012). *Auditory Comprehension: intact and impaired.* Presented at Chest Heart and Stroke Study Day: “Technology”; 2012 November 21; Edinburgh, UK.

Poster presentations:

Jones, A. (2013). *Computer-assisted Speech and Language Therapy: preliminary results from auditory comprehension therapy in aphasia.* Presented at Biennial Meeting of the World Federation of Neurology Research Group on Aphasia, Dementia and Cognitive Disorders; 2013 December 12; Hyderabad, India.

Jones, A. (2012). *Computer-assisted home speech and language therapy in aphasia.* Presented at SINAPSE Annual Scientific Conference; 2012 May 30; Glasgow, UK.

Jones, A. (2011). *Investigating the benefits of computer-assisted home speech and language therapy in Wernicke aphasia.* Presented at SINAPSE Annual Scientific Conference; 2011 June 16; Dundee, UK.

Contributions by Anna Jones: study design, ethical submission, patient recruitment, data collection and analysis, results interpretation, writing up of the study, writing and giving oral and poster presentations.

Word/picture association tests are used in many neuropsychological tests, including (but not limited to): Pyramids and Palmtrees (PPT) (Howard and Patterson, 1992); Test for Auditory Comprehension of Language (TACL) (Carrow-Woolfolk, 2014); Peabody Picture Vocabulary Test-4 (PPVT-4) (Dunn and Dunn, 2007); as well as sections of the Western Aphasia Battery (WAB) (Kertesz, 1982), the Boston Diagnostic Aphasia Examination (BDAE) (Goodglass et al., 2001) and the Psycholinguistic Assessments of Language Processing in Aphasia (PALPA) (Kay et al., 1992). As argued by Hillis and collaborators (2004), testing this association using a single choice is however not very sensitive (50% chance level). Since only one previous study has looked at the word/picture verification task employed in Chapter 6 in the clinical population (Breese and Hillis, 2004), it is important to perform a comparable study. The Breese and Hillis study demonstrates the high sensitivity of this task in identifying auditory comprehension impairments and discusses the importance of error pattern investigation. However the actual quality of the task was not investigated in terms of correlations with behavioural or neural clinical presentations of patients. Thus, this study aims to explore the task quality in more detail.

Secondly, Breese and Hillis's study (2004) supports the distributed theory of semantic representation over other local models. One of the aims of the research presented here is to extend these findings by further investigating the cognitive neuropsychological model of single word processing presented in Chapter 2 (Figure 2-8– extended version of Whitworth's model (Whitworth et al., 2014)), with regards to the picture/word verification task. Based on the fMRI results in Chapter 6 showing incongruency effects, we expect that a lesion within the left pars triangularis will result in impaired phonological auditory processing, whereas a lesion within the left angular gyrus will result in impaired semantic auditory processing.

Thirdly, the use of the task to detect changes over time in auditory comprehension has not been investigated. Therefore, this task was used to investigate changes in auditory comprehension skills in a small subgroup of participants carrying out an independent computer-assisted therapy, specifically focussing on auditory comprehension impairments. The primary aim of speech and language therapy in aphasia is to maximise communication ability and, although there is no universally accepted treatment for all patients, there are a variety of approaches that are commonly employed by speech and language therapists. The main two methods used are: impairment-based therapy and communication-based therapy/ social-participation approaches. Impairment-based therapy focusses on improving language functions, whereas communication-based therapy aims to enhance communication by any means, including increasing social participation. Impairment-based therapy for aphasia was first formally introduced in 1930s (Weisenberg and McBride, 1935) and has been heavily influenced by models of disability, functional neuroanatomical models and cognitive neuropsychological models. The effectiveness of aphasia treatment is still highly debated, although recent meta-analyses and reviews suggest that impairment-based therapy is beneficial for aphasia (Brady et al., 2012; Kelly et al., 2010; Robey, 1998). The approach presented here therefore focuses on treating impaired cognitive functions.

Recovery from aphasia, both spontaneous and treatment-based, is supported by the functional and structural neuroplasticity of the brain. The exact mechanisms underlying aphasia recovery are still debated, but spontaneous language recovery has been associated with a reactivation of the peri-lesioned left perisylvian cortex (Heiss and Thiel, 2006) after an early upregulation by the right hemisphere (Saur et al., 2006), thus supporting the dual-stream model of processing (Chapter 1, Figure 1-6 (Rauschecker, 2011)). As speech and language therapy aims to stimulate the cortical networks involved in language and maximise neuroplasticity, the timing of therapy delivery is an important debate. It has been shown that intensive impairment-based language therapy improve chronic stages of aphasia (Bhogal et al., 2003), even when only given for short period of time (Meinzer et al., 2005). However, it is also known that half of stroke recovery occurs within the first month and occurs for up to 6 months post stroke (Wade et al., 1986), with the majority of spontaneous recovery occurring in the first 3 months following the acute phase i.e. ‘critical period’ (Alexander, 1994; Mazzocchi and Vignolo, 1979). Therefore, an optimal regime of early implemented impairment-based therapy for aphasia following stroke is currently held (Godecke et al., 2012; Robey, 1998; Teasell et al., 2005). Several studies have looked at the intensity of speech and language therapy for aphasia and again, although no definitive quantitative conclusion has been made, it is suggested that intensive therapy can be effective (Bhogal et al., 2003). However, limited resources (for example, financial time pressures) are prohibitive to the intensive input that is required (Palmer et al., 2012). Although impairment-based therapy can be delivered through group treatment, it is most commonly administered through didactic interactions between the speech and language therapist and the patient. However, computerised aphasia therapy has been reported to be valuable in providing targeted impairment-based practice at higher intensities, with many improved outcomes (Cherney, 2010; Fink et al., 2002; Mortley et al., 2004; Palmer et al., 2012; Sandt-Koenderman, 2011). There are a number of challenges to using a computer for language rehabilitation, including cost of and access to resources, lack of computer literacy, motor impairments, cognitive impairments, and lack of support. However, if these issues are addressed appropriately and patients are provided with the right level of support, the delivery of computer therapy for aphasia post-stroke has been shown to be successful in terms of feasibility, cost and quality of life (Latimer et al., 2013; Palmer et al., 2012). The approach presented here therefore focused on delivering a computer-based therapy to perform at home (independently) as early as was feasible following stroke, and as intensively as possible.

7.1.2. Method

Ethics Statement

This study was approved by the NHS Scotland A Research Ethics Committee (*REC reference number: 11/AL/0069*), NHS Lothian Research and Development (*R&D project number: 2011/W/ST/01)*, Edinburgh Clinical Research Facility (*CRF approval number: E11928 CRF*) and the University of Edinburgh Psychology Department (*Psychology Research Ethics number: 02-1112*). All participants were provided with detailed information regarding the nature of the study and their potential role in it and full informed consent was obtained from participants. All information was provided in an aphasia-friendly manner, and all consent taken by a qualified and experienced Speech and Language Therapist (see Appendix 17 and 18). All participants agreed to their G.P. receiving a letter outlining the study and explaining participation (see Appendix 19).

Participants

All patients were recruited into the study from 5 sites across NHS Lothian stroke services (4 inpatient and 1 outpatient site). All healthy participants were recruited through the patient cohort (i.e. relatives) or the University of Edinburgh's Psychology Department volunteer panel. All participants were right-handed, as assessed using the Edinburgh Handedness Inventory (Oldfield, 1971), were native English speakers with no uncorrected hearing or visual impairments, and had sufficient dexterity control to carry out scanner responses. Healthy participants were excluded if they had any known neurological disease/ disorder and/or a communication impairment of any kind. Patients were recruited according to their diagnostic NHS CT and/or MRI scan and their behavioural profile - all patients had a left hemisphere frontal, temporal and/or parietal stroke lesion (ischaemic or haemorrhagic) visible on their scan, with an associated clinically assessed aphasia, and had no prior diagnosis of, nor presented with symptoms of, a dementia. Patients were all recruited into the study a maximum of 5 weeks post-stroke. Importantly, only individuals who were able to make informed consent were eligible to participate in this study.

In total, 21 participants were recruited into the study (15 stroke patients: 10 male, 5 female, 63.4 +/- 14.5 years; 6 healthy controls: 2 male, 4 female, 56.3 +/- 15.12 years).

Procedure

Time point 1 (5-6 weeks post-stroke)

All participants (stroke patients and healthy controls) were involved in the first time point of the protocol, including behavioural assessments and structural/ functional brain scanning (Figure 7-1). In the analysis, only behavioural and structural imaging were investigated.

1. Behavioural assessment

A battery of behavioural (language and cognitive) assessments was carried out with each participant over a maximum of 2 sessions.

Language assessments included: The Western Aphasia Battery (language component) (Kertesz, 1982) to diagnose aphasia and investigate severity/type; the Peabody Picture Vocabulary Test 4 (PPVT: 4) (Dunn and Dunn, 2007) to investigate single word auditory comprehension; and four short experimental assessments to investigate semantics: 1. PPT: sub-sections from The Pyramids and Palmtrees Test (nouns) (Howard and Patterson, 1992); 2. KDT: sub-sections from The Kissing and Dancing Test (verbs) (Bak and Hodges, 2003); 3. TTT: Tuna and Tomatoes (sequencing); and 4. SPMT: sound to picture matching test (non-verbal) (Danek et al., 2013).

Cognitive assessments included: Depression Intensity Scale Circles (DISCS) (Turner-Stokes et al., 2005) and Visual Analogue Self-esteem Scale (VASES) (Brumfitt and Sheeran, 1999), to screen for those with low mood; Bell Test (Gauthier et al., 1989) to screen for hemianopia and impaired non-verbal attention; Wechsler Adult Intelligence Scale-III (WAIS III): digit span (Ryan and Lopez, 2001) to assess verbal short-term memory and working memory; Doors and People Test: visual recognition (Baddeley et al., 1994), to investigate non-verbal short-term memory.

All participants, as well as Speech and Language Therapists and G.P's if appropriate, were provided with their full assessment results if requested (Appendix 20).

2. Imaging

All participants also underwent structural and functional imaging at time point 1. Imaging was obtained using a GE Signa HDxt 1.5T scanner at the Brain Research Imaging Centre (http://www.bric.ed.ac.uk/), University of Edinburgh, UK. Structural imaging consisted of a T1-weighted volume (3D IRP - 160 slices, 1.3 mm thick coronal slices, 1.3 x 1.3 mm in-plane resolution with a 256 mm FOV), a T2-weighted volume (EDR FC – 37 slices, 4 mm thick axial slices, 1 x 1 mm in-plane resolution with a 256 mm FOV). Functional MRI was performed to map single word auditory comprehension areas (EPI – 30 slices, 4mm thick axial slices, 4 x 4 mm in-plane resolution with a 256 mm FOV, TR = 2500 msec, TE = 50 msec). The task was identical to that presented in Chapter 6. Resting state fMRI (5 mins) was also performed, using the same imaging parameters as for the active state.

3. Group allocation

The patient cohort was randomly split into two groups: Group A - Therapy group (n= 6; 4 male, 2 female; 56.5 +/- 13.0 years); Group B – No Therapy group (n= 9; 6 male, 3 female; 68 +/- 14.2 years).

Time point 2 (18-19 weeks post-stroke)

At time point 2, all participants underwent the same battery of behavioural assessments and the same imaging (structural and functional) protocol as at time point 1 (Figure 7-1). Different versions of the PPVT 4 and the fMRI task (word/picture verification) were administered to time point 1, to counteract for any test-retest reliability issues.

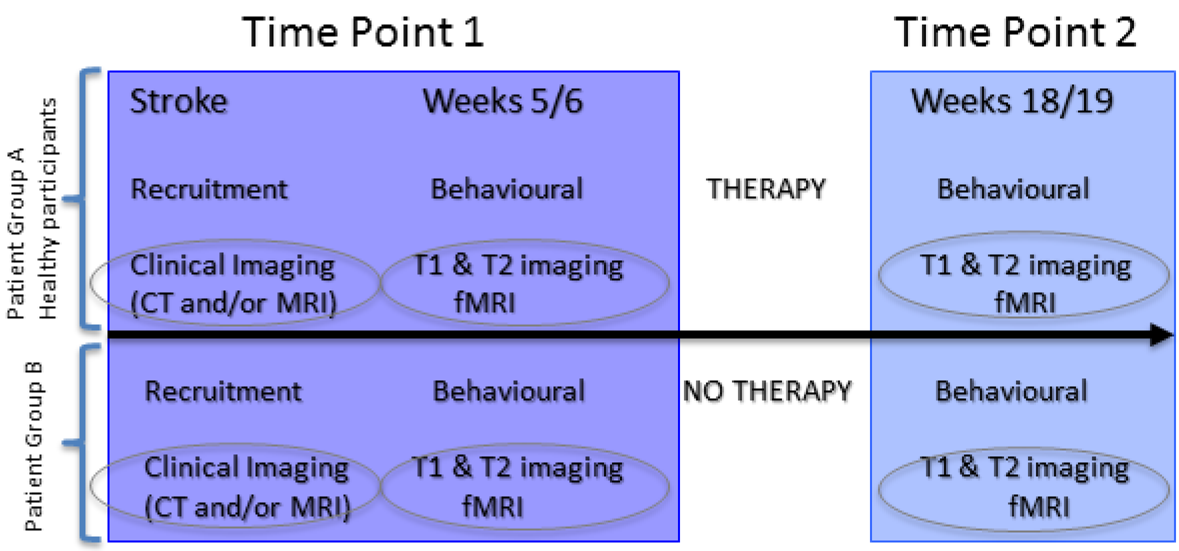

Figure 7-1: Protocol for participant groups (2 patient groups and 1 healthy participant group).

Feedback forms were given to all participants who carried out the therapy, at time point 2 (Appendix 21).

Therapy

Participants randomised to patient group B continued usual care i.e. any NHS SLT intervention, as well as daily life conversation, reading and writing activities, and attendance at any support groups that they were already involved in. Participants randomised to patient group A also continued to participate in their usual activities, as above. In addition, this group as well as the healthy control group, received speech and language therapy intervention through an independent computer therapy programme called React2© (http://www.react2.com/) between weeks 5/6 and 18/19 (i.e. a period of approximately 3 months). Specific modules/levels from the React2© programme were chosen, tailored according to individual abilities, configured by an SLT (Anna Jones) and provided on a tablet computer.

The React2© computer programme consists of speech and language therapy exercises. The modules chosen were all designed to focus on auditory comprehension impairments at the level of single words (nouns/verbs) and sentences, as well as working on semantics (Table 7-1). A variety of React2© exercises were provided to each individual including matching auditory presented single nouns/verbs to pictures (varies by degree of semantic relatedness); matching auditory presented short phrases/ complex sentences to pictures (varies by sentence complexity and semantic relatedness); and classification/categorisation (or elimination) of auditory or visually presented single nouns/verbs (varies by semantic relatedness) (Table 7-1, Figure 7-2). Importantly, no spoken or written output was required by participants, and all information was provided aurally and/or visually. Records of length of time, frequency of use and scores were recorded automatically by the programme.

Main module	**Sub-module**	**Tasks**	**Levels within sub-module**	**"Instructions" / Description**
Auditory Processing	Understanding words	Pointing to a noun by name	7: increase in number of items and semantic relationships	*"listen to a noun, then click on the picture which matches"*
		Pointing to a verb by name	3: increase in number of items and semantic relationships	*"listen to a verb, then click on the picture which matches"*
	Understanding sentences	Pointing to a picture by description	5: increase in grammar complexity	*"listen to a sentence, then click on the picture that matches"*
		Sentence judgement – semantics	3: increase in grammar and semantic complexity	*"listen to a sentence, then decide if it is correct or incorrect"*
Semantics	Classify/ categorise	Classify/ categorise pictures	3: increase in number of items.	*"click on the play button, then click on what it tells you"*
		Classify/ categorise words	3: increase in number of items.	*"click on the play button, then click on what it tells you"*
	Category elimination	Eliminating pictures	2: increase in number of items	*"listen to a phrase, then click on the picture which matches"* *- negation pictures*
		Eliminating words	2: increase in number of items	*"listen to a phrase, then click on the word which matches"* *- negation words*

Table 7-1: Modules, sub-modules and tasks from React2© focussing on auditory comprehension, showing number of levels present, type of variance between modules and instructions for participants.

Initial tuition was provided to both the participant and carer on how to use the tablet computer and how to work through the exercises. Each intervention participant was contacted a minimum of every 10 days with the aims of helping solve any difficulties, facilitating participation and sustaining motivation. Participants were advised to use the computer exercise for a minimum of 20 minutes, 5 days a week for 3 months. No maximum time was advised.

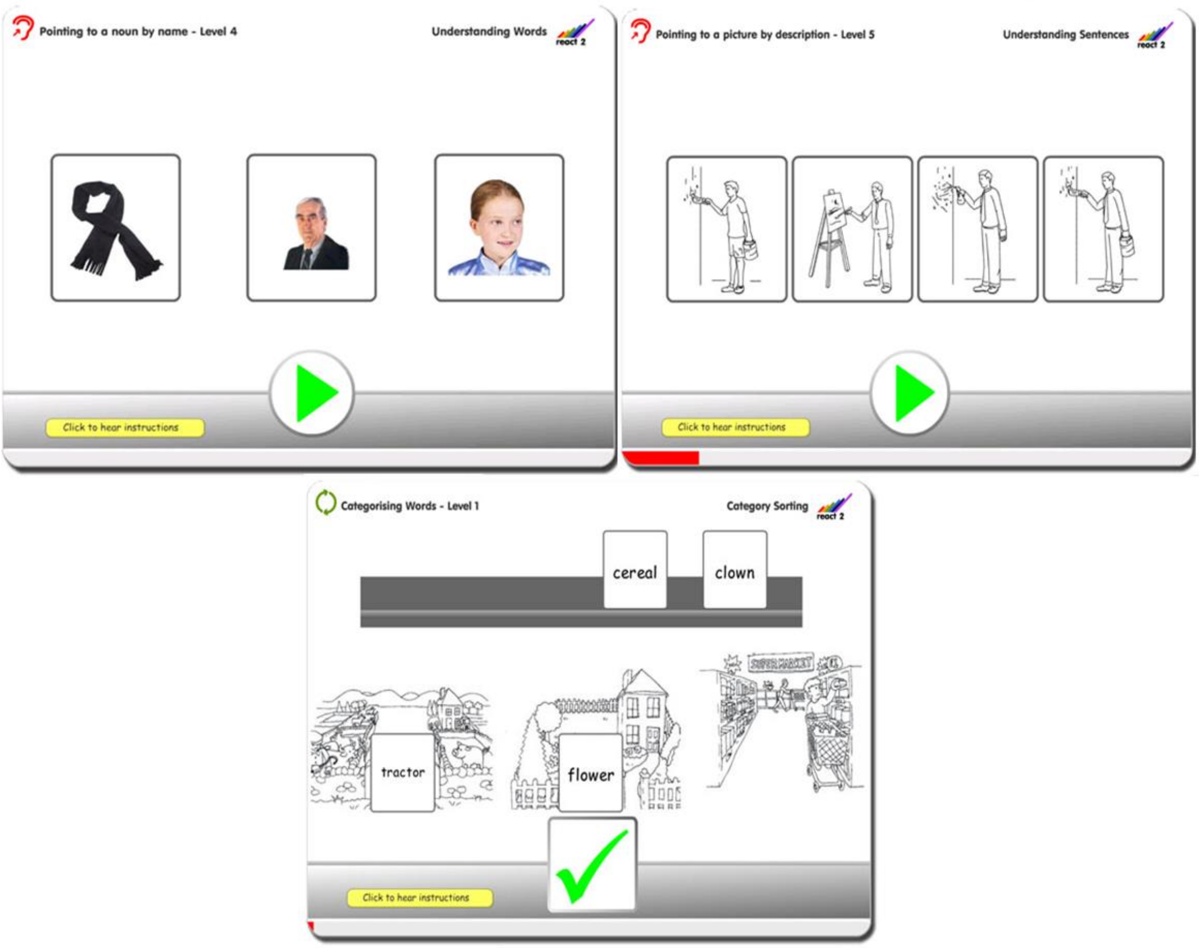

Figure 7-2: Example of exercises from React2© focussing on auditory comprehension/ semantics. *Top left:* Matching auditory and visual noun, level 4. *Top right*: Matching auditory sentence with pictorial representation, level 5. *Bottom*: Putting nouns into categories, level 1.

7.1.3. Data Analysis

Picture/ word verification task analysis – time 1

For the patient population at time point 1, overall performances on the picture/word verification task for each subject were calculated out of 30 (i.e. to score 1 point the participant had to correctly answer the target that matched and all the foils that did not match). Scores (%) and reaction times (RTs – sec) for each condition (correct/ phonological/ semantic/ unrelated) were calculated, per participant. Any incongruency effects (i.e. matched conditions vs. unmatched conditions) and differences between the incongruency effects (i.e. differences between the unmatched conditions) were investigated by performing multiple comparisons between data pairs (percentile bootstrap on differences).

A robust correlation analysis was performed to investigate any correlation between performance on the task and on the following language and cognitive assessments: Western Aphasia Battery (WAB) comprehension section (Kertesz, 1982); Peabody Picture Vocabulary Test 4 (PPVT-4) (Dunn and Dunn, 2007), Pyramids and Palmtrees Test (PPT) (Howard and Patterson, 1992), Kissing and Dancing Test (KDT) (Bak and Hodges, 2003), Wechsler Adult Intelligence Scale-III (WAIS III): Digit Span (Ryan and Lopez, 2001) and Doors and People Test: visual recognition (Doors) (Baddeley et al., 1994). These correlations were investigated because it can be expected that there will be a link between the semantic performances measured by the word/picture verification task (Breese and Hillis, 2004) and the overall comprehension (and thus the semantic component) of the WAB, as well as with the PPT and KDT performances, as all of these investigate semantic processing and underlying semantic knowledge. It can also be expected that there will be a link between the auditory comprehension of single nouns measured by the word/ picture verification task and the PPVT4, as well as the comprehension component of the WAB, as these all investigate single word auditory comprehension. Finally, it can also be expected that there will be a link between performance on the word/ picture verification task and performance on the Digit Span and the Doors subtests as these assessments measure both verbal and visual short-term memory - skills that are required when carrying out the word/picture verification task.

Lesion location analysis – time 1

Preprocessing of structural scans

Lesion masks were created by manually segmenting the lesions visible on the T2 structural images using MRIcron (Rorden, 2007), and coregistered to the T1 image using normalized mutual information, using SPM12 (Wellcome Trust Centre for

Neuroimaging, London, UK, http://www.fil.ion.ucl.ac.uk/spm/). The T1 images were normalized to standard space using SPM12 and the derived wrapping field applied to the lesion masks. Lesion locations were next identified in SPM using the Anatomy Toolbox (Eickhoff et al., 2005) and the WFU PickAtlas (Wake Forest University, Winston-Salem, NC, http://www.ansir.wfubmc.edu/). Eight Regions of Interest (ROIs) were chosen to classify lesions: pars triangularis, left insula, Wernicke's area, left angular gyrus, left precuneus, left temporal pole, primary auditory cortex and the motor cortex (primary motor cortex, premotor cortex and supplementary motor area). Wernicke's area was determined using a composite definition involving the angular gyrus, supramarginal gyrus, and posterior sections of the superior, middle and inferior temporal gyri (Mesulam et al., 2015).

Behavioural analysis

Patients were differentiated according to those that performed above and those that performed below chance level (using binomial confidence intervals: 31.3-68.7%) on incongruent conditions on the picture/word verification task, and grouped according to lesion location. Group comparisons (according to performance on task) were investigated by performing multiple comparisons between data pairs (percentile bootstrap on differences). Average performances across lesion locations were also compared.

Therapy analysis – time 2

As data was automatically stored on the participant tablet computers, it was possible to calculate the following measures for the 3 month therapy period: average score achieved on therapy (%), total time (mins) spent on therapy, the average time spent per task (mins), average number of exercises completed, and total number of exercises completed. The average time spent on therapy per day (min) was also calculated. The effect of the computer-assisted React2© therapy was investigated by comparing changes (time 2 - time 1) on the task and other behavioural measures (language and cognitive) with therapy measures. Spearman correlations were performed, without correction for multiple comparisons, as a purely exploratory analysis. The mean differences between the different participant groups (healthy, patient group A and patient group B) were tested, with an adjustment for multiple comparisons, to compare differences in behavioural score changes across time 1/time2.

7.1.4. Results

Picture/ word verification task analysis – time 1

Performance on the auditory comprehension task, for patients with a diagnosed post-stroke aphasia at the acute phase (time point 1), was poor, with participants attaining an average overall performance score of 6.47 +/- 7.06 (max. 30). Both accuracy rates and reaction times showed a large variance per condition. Accuracy rates per condition ranged from an average of 80.63% +/- 25.87 (correct) to 50.73% +/- 37.17 (semantic), and reaction times per condition ranged from 1.58 sec +/- 0.45 (correct) to 1.86 sec +/- 0.51 (semantic) (Table 7-2, Figure 7-3).

	Overall performance (max. 30)	**Accuracy rates (%)**				**Reaction times (sec)**			
		C	**P**	**S**	**U**	**C**	**P**	**S**	**U**
mean	6.47	80.63	53.49	50.73	64.69	1.58	1.76	1.86	1.75
S.D.	7.06	25.87	32.04	37.17	41.74	0.45	0.39	0.51	0.56

Table 7-2: Mean (with standard deviations) overall performance and accuracy rates (%)/ reaction times (sec) per condition.
C: correct target; P: phonological foil; S: semantic foil; U: unrelated foil.

Performance on the auditory comprehension task for the patient population at time point 1 showed an incongruency effect for the semantic condition (-29.9%, p=0.020) and no significant difference between the correct condition and the phonological (-27.14%, p=0.024) or unrelated foils (-15.94%, p=0.928). In terms of reaction times there was a significant incongruency effect for the phonological conditions (p=0) and the semantic conditions (p=0). However, no significant differences were seen between the incongruent effects in terms of reaction times (Table 7-3, Figure 7-3).

	Accuracy rate		
Condition pairs	**C/P**	**C/S**	**C/U**
Difference	-26.2	-21.3	0.65
CI	[-68 4]	[-66 0]	[-58 14]
p-value	0.024	0.020	0.928
	Reaction times		
Condition pairs	**C/P**	**C/S**	**C/U**
Difference	*26.5*	*27.3*	17.1
CI	*[6 45]*	*[9 43]*	[-5 43]
p-value	*0*	*0*	0.086
	Reaction times		
Condition pairs	**P/S**	**P/U**	**S/U**
Difference	0.5	5.9	11.8
CI	[-14 12]	[-9 29]	[-9 28]
p-value	0.960	0.410	0.120

Table 7-3: Percentile bootstrapped pair-wise differences on accuracy rates and reaction times between matched conditions and unmatched conditions.

C: correct target condition; P: phonological foil condition; S: semantic foil condition; U: unrelated foil condition. Significant differences shown in italics.

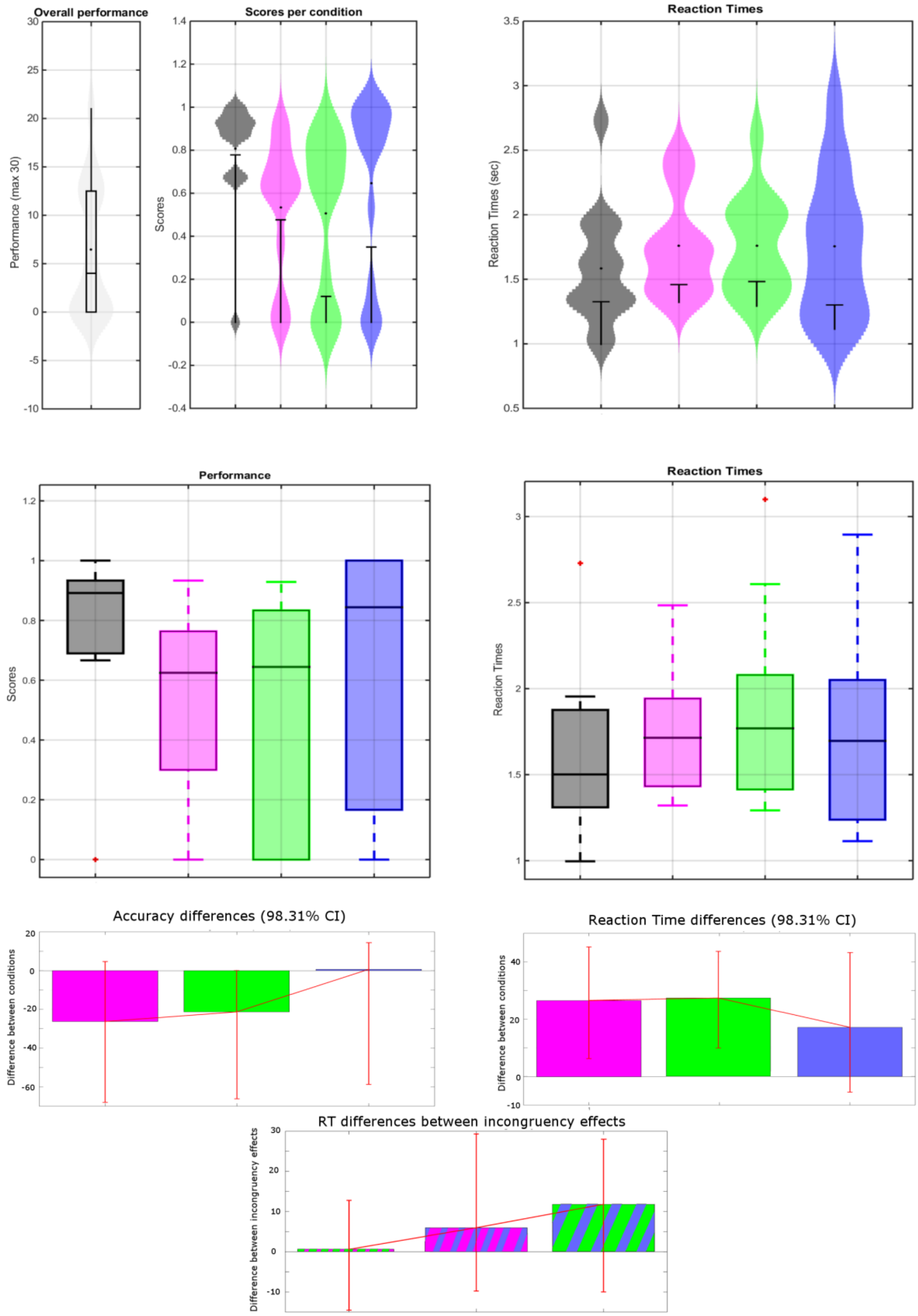

Figure 7-3: Behavioural results for fMRI auditory comprehension task for all patients at time point 1.

Colours show 4 conditions: grey- correct target; pink- phonological foil; green- semantic foil; blue- unrelated foil.

Row 1: Violin plots (boxplot showing IQ range with a rotated kernel density plot on each side to signify the probability density of the data at the different values (Hintze and Nelson, 1998)) showing behavioural results per condition for (left – right) overall score (max 30), performance per condition and reactions times (sec), with the median data point (black dot) overlaid.

Row 2: Box plots showing behavioural results per condition for (left-right) performance per condition and reaction times (sec), with outliers shown (red dot).

Row 3: Pair-wise comparison graphs showing the differences between pairs of conditions with confidence intervals shown (red – 98.31%) for (left-right):

- accuracy rates for correct (congruent) vs. incongruent conditions (phonological – semantic – unrelated);
- reaction times for correct (congruent) vs. incongruent conditions (phonological – semantic – unrelated)

Row 4: Pair-wise comparison graph showing the differences between incongruent effects (phonological vs. semantic – phonological vs. unrelated -semantic vs. unrelated) with confidence intervals shown (red – 98.31%)

Robust correlation analysis showed a weak, non-significant, correlation between the patient scores on the word/picture task and scores on the comprehension section of the WAB, r=0.6 [-0.2 0.9], and a significant correlation between the patient task scores and scores on the Doors subtest, r=0.8 [0.3 0.9] (Figure 7-4). None of the other behavioural scores correlated with the scores on the word/picture task for patients at time 1.

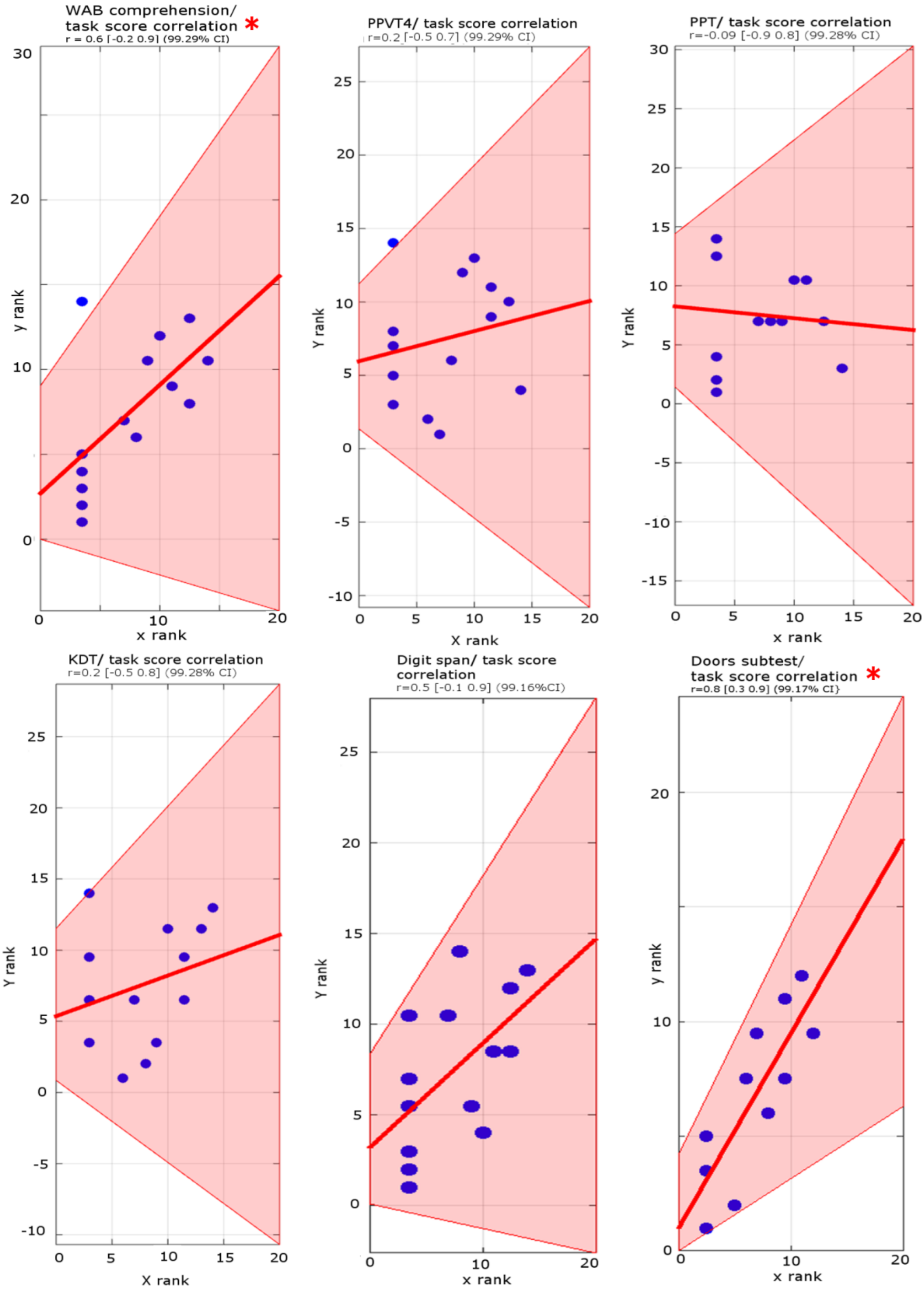

Figure 7-4: Robust Spearman correlation scores for the 6 main assessment scores, with the task score in the patient population at time 1.

Top row (left-right): WAB, PPVT4 and PPT score correlations with the task score.

Bottom row (left-right): KDT, Digit span and Doors subtest score correlations with the task score.

Red asterisk signifies correlations of note: Doors subtest showed a significant correlation and the WAB comprehension test showed a correlation (non-significant) with the task score.

Lesion location analysis – time 1

In terms of fronto-temporal lesions, 87% of patients showed damage in the left insula, 53% in the pars triangularis of the left inferior frontal gyrus (Brodmann area 45 – part of Broca's area), 46% in the primary auditory cortex, 40% in the left temporal pole, and 33% in Wernicke's area (Mesulam et al., 2015). 13% of patients showed damage to the motor cortex. In terms of the parietal lobe, 20% showed damage in the left precuneus and 13% in the left angular gyrus (Table 7-4).

Patient	**Pars triangularis**	**Left insula**	**Wernicke's area (composite)**	**Left temporal pole**	**Primary auditory cortex**	**Angular gyrus**	**Left precuneus**	**Motor cortex**
1	X	X		X				
2		X						
3	X	X		X	X			X
4	X	X	X	X	X			
5		X					X	
6		X						
7			X			X		
8		X			X			
9		X						
10	X	X		X	X			X
12	X	X			X			
13	X	X	X	X	X			
11	X	X	X	X	X	X	X	
14			X				X	
15	X	X						
TOTAL	**8**	**13**	**5**	**6**	**7**	**2**	**3**	**2**

Table 7-4: Lesion distribution, according to the 8 defined ROIs, for all patients (n=15) as seen on T2 scans at time point 1.

Green: performed above chance level for all incongruent conditions on the picture/word verification task. *Red*: performed below chance level for all incongruent conditions on the picture/word verification task. *Blue*: performed below chance level only on the phonological condition on the picture/word verification task. *Note:* participant number 7 was unable to perform the task and so did not receive a score for any of the conditions.

20% of patients (3/15) performed above chance level for all incongruent conditions on the word/ picture verification task, and 20% of the patients (3/15) performed below chance level for all incongruent conditions. Multiple group comparisons showed significant differences for all conditions between these two patient groups: correct condition group difference -0.18 [-0.27 -0.09]; phonological condition group difference 0.84 [0.75 0.92]; semantic condition group difference 0.87 [0.81 0.92]; unrelated condition group difference 0.92 [0.86 0.98].

Among patients who performed above chance level, 1 had a lesion in the left pars triangularis, and none had damage to the precuneus or angular gyrus. Among patients who performed below chance level for all conditions, all (3/3) had damage in the left pars triangularis, and 1 also had damage to the precuneus and the angular gyrus
Table 7-4 & Figure 7-5).

13% patients (2/15) performed above chance level for the correct, semantic and unrelated conditions but below chance level for the phonological condition (Table 7-4). 1 of these patients showed damage to the pars triangularis and the other to the precuneus, and neither showed any damage to the angular gyrus. No patients performed above chance level for the correct and phonological conditions and below chance level for the semantic and unrelated conditions (Figure 7-5).

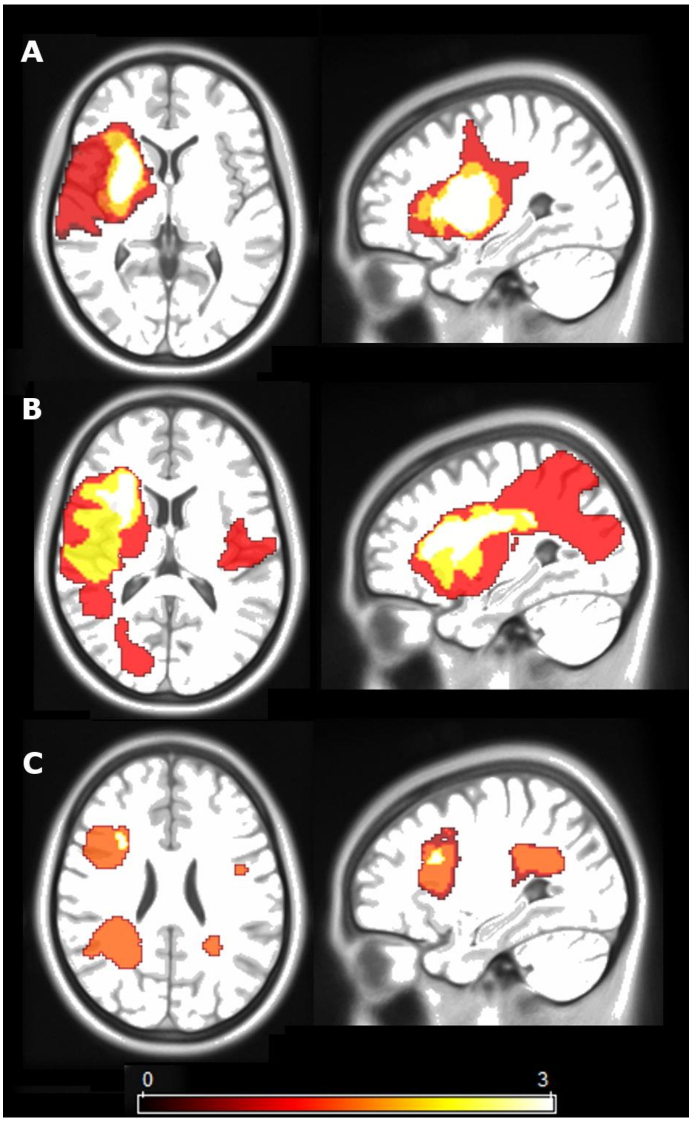

Figure 7-5: Lesion distribution for 3 patient groups:

A) n=3 patients who performed above chance level for all incongruent conditions on the picture/word verification task; B) n=3 patients who performed below chance level for all incongruent conditions on the picture/word verification task; C) n=2 patients who performed below chance level only on the phonological condition on the picture/word verification task. Red: 1 patient; yellow/orange: 2 patients; white: 3 patients. Lesion masks obtained from MNI aligned T2 images and overlaid onto the ICBM-152 T1 MNI template (Mazziotta et al., 2001).

The average task performance (accuracy rate) in patients with damage to the angular gyrus (N=2) was 0 (note one patient was unable to perform task and so did not respond), whereas the average performance in those with damage to the pars triangularis (N=8) was 0.39 (Figure 7-6).

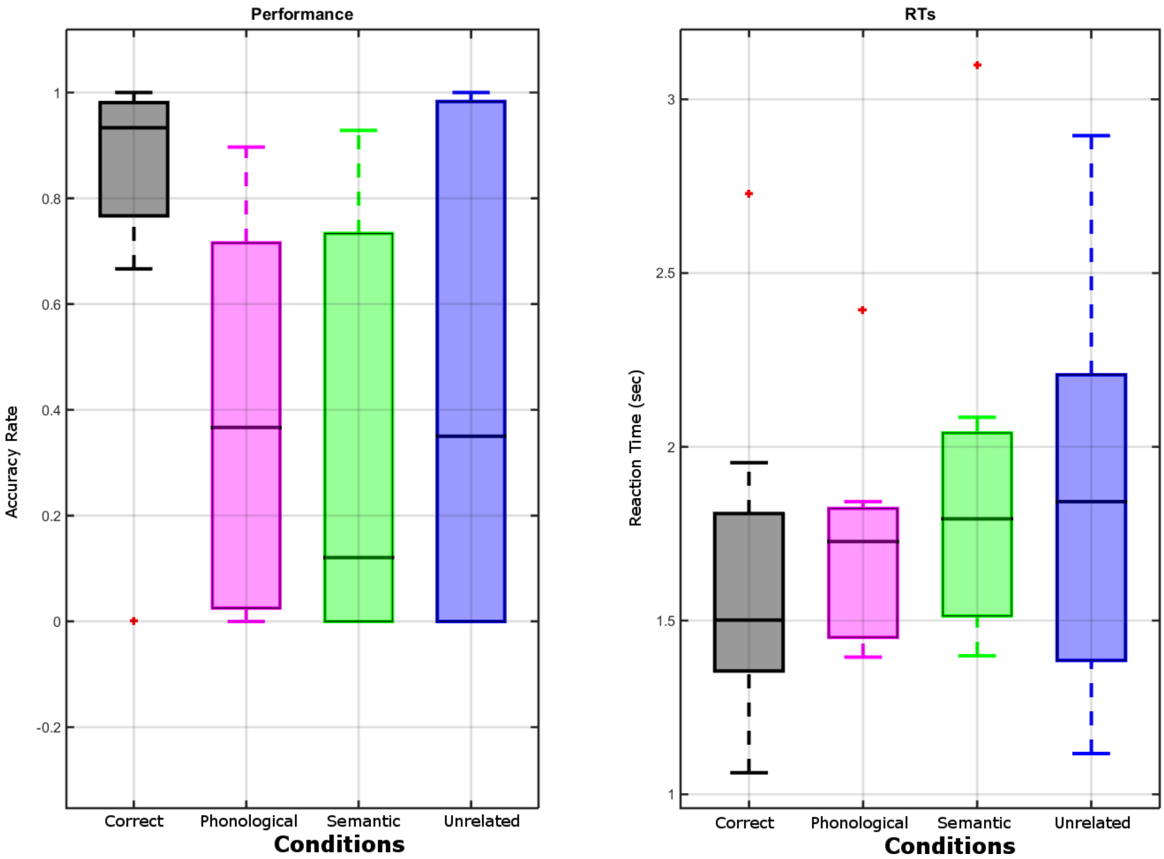

Figure 7-6: Boxplots showing performance and reaction times for patients with damage to the left pars triangularis (n=8), for each condition.

Colours correspond to conditions - grey: correct; pink: phonological; green: semantic; blue: unrelated. Red crosses signify outliers. Performance and reaction times for patients with damage to the left angular gyrus (n=2) not shown as 1 patient was unable to complete task and therefore his scores/ RTs cannot be plotted.

Therapy analysis – time 2

At time point 2, all healthy participants (n=6) and all patients in Group A (n=6) returned for further assessments. Five individuals from patient Group B withdrew from the study after time point 1, thus patient group B totalled n=4 for any analysis included at time point 2.

Therapy measurements calculated included mean number of days spent on therapy, mean time (minutes) spent on therapy, mean number of exercises completed, average score on therapy over the 3 month period (%) and mean time spent on therapy per task (min) and per day (min) (Table 7-5 & Figure 7-5).

	Patient Group (n=6)	**Healthy Group (n=6)**
Average number of days on therapy	60.33 (+/- 45.23)	78.83 (+/- 19.43)
Average time on therapy (min)	1837.80 (+/- 1356.00)	1437.61 (+/- 944.40)
Average number of exercises completed	463.5 (+/-399.59)	362.67 (+/-265.51)
Average time spent per day (min)	27.62 (+/-5.23)	18.62 (+/- 8.07)
Average time spent per task (min)	6.18 (+/- 3.50)	4.10 (+/- 2.46)
Average score (%)	99.14 (+/-0.84)	99.07 (+/-1.53)

Table 7-5: Therapy data for patient group A and the healthy control group

Differences between the scores at time point 1 and time point 2 were calculated for the task (Figure 7-7, top row) and the main language/ cognitive assessments. Changes in therapy measures were correlated with changes in behavioural and task scores. No significant correlations were found between the task behavioural score differences and the amount of time spent on therapy (min), the average score achieved on the therapy (%), or the number of exercises carried out for the therapy. No significant correlations were found between the same therapy measures and the difference in scores (time 2-time1) for the Peabody Picture Vocabulary Test 4 (PPVT: 4) (Dunn and Dunn, 2007), The Kissing and Dancing Test (verbs- KDT) (Bak and Hodges, 2003), the Wechsler Adult Intelligence Scale-III (WAIS III): digit span (Ryan and Lopez, 2001), nor the Doors and People Test: visual recognition (Baddeley et al., 1994). A significant correlation was found between the number of therapy exercises carried out and the change in participant scores on the The Pyramids and Palmtrees Test (nouns- PPT) (Howard and Patterson, 1992) ($r=0.9$ [0.6 1.0], $p=0.008$). Trends (considered here as $r>=0.7$ showing >49% variance, i.e. the coefficient of determination) were found between 1. the average therapy score (%) and the change in participant scores on the comprehension subtest of the Western Aphasia Battery (WAB) (Kertesz, 1982) ($r= 0.7$ [-0.9 1.0], $p=0.117$) and 2. the total time spent on therapy (min) and the change in participant scores on The Pyramids and Palmtrees Test ($r=0.77$ [-0.5 1.0], $p=0.07$) (Figure 7-7, middle row).

Multiple group comparisons revealed a weak difference (not significant) between patient group A (therapy) and B (no therapy) on the comprehension subtest of the Western Aphasia Battery (1.08 [-0.03 -5.72]) and on The Pyramids and Palmtrees Test

(-2.42 [2.20 0.88]). A significant difference was found between patient group A (therapy) and the healthy group on the comprehension subtest of the Western Aphasia Battery (2.11 [1.15 -2.31]), and a weak different (not significant) on The Pyramids and Palmtrees Test (-1.00 [3.06 0.31]) (Figure 7-7, bottom row).

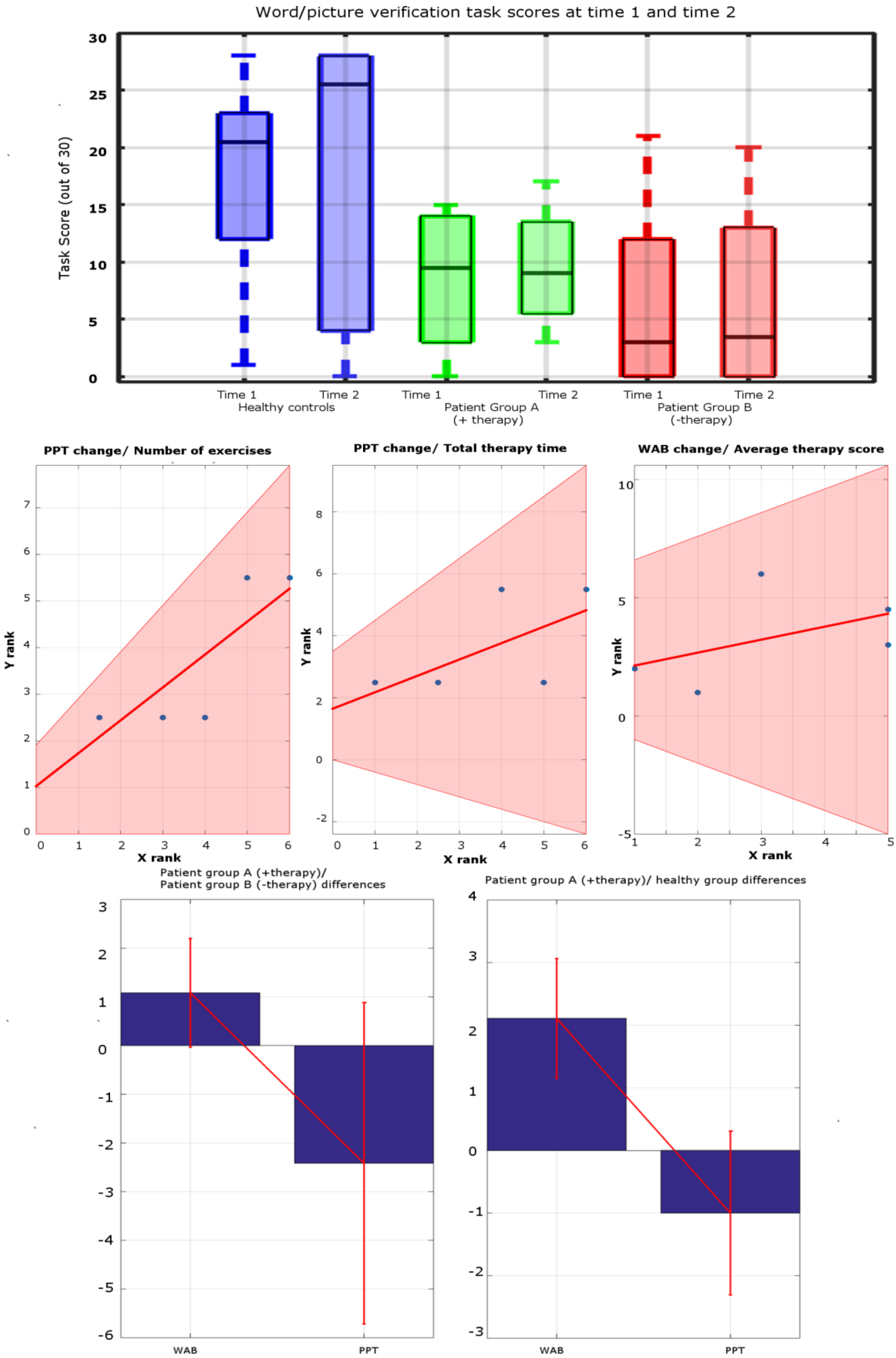

Figure 7-7: Time 2 scores and correlations.

Top row: Boxplots showing time 1 and time 2 global scores on the word/ picture verification task for healthy controls (blue, n=6), patient group A (+ therapy, green, n=4), and patient group B (-therapy, red, n=6). *Middle row*: Spearman correlations between language score change and therapy measures. Left – right: change in Pyramids and Palmtrees Test score significantly correlates with the number of exercises participants carried out (r=0.9 [0.6 1.0], p=0.008); change in Pyramids and Palmtrees Test score shows a trend correlation with the total time participants spent on the therapy (r= 0.7 [-0.9 1.0], p=0.117); change in Western Aphasia Battery comprehension score shows a trend correlation with the average time spent on therapy (r=0.77 [-0.5 1.0], p=0.07). *Bottom row*: Multiple group comparisons between (left) patient group A (with therapy)/ patient group B (without therapy) and (right) patient group A (with therapy)/healthy group on the changes in the Western Aphasia Battery comprehension (WAB) and the Pyramids and Palmtrees Test (PPT) scores between time 1 and time 2. The WAB comprehension score difference is significantly different (difference = 1.08 [-0.03 -5.72]) between patient group A (with therapy) and the healthy group.

Feedback on React2© therapy

100% of participants who carried out the therapy returned feedback sheets, and all patients reported that they ‘felt the React2 therapy was helpful’. The degree to which patients reported the therapy was helpful or not, according to the Likert Scale (Likert, 1932) (0-5 / not helpful-very helpful) in specific areas of communication, varied (Figure 7-6). 83% patients reported that the therapy was ‘helpful’ (4) or ‘very helpful’ (5) with their auditory comprehension abilities ($\widetilde{x}$ = 4, IQR = 0.75); 33% with their reading abilities ($\widetilde{x}$ = 3, IQR = 2.5); 83% with their spoken output ($\widetilde{x}$ = 4, IQR = 0.75); and 100% with their memory ($\widetilde{x}$ = 4, IQR = 0.75) (Table 7-6).

	Auditory comprehension	**Reading**	**Spoken output**	**Memory**
Participant 1	0	0	5	5
Participant 2	4	4	4	4
Participant 3	4	3	4	4
Participant 4	4	3	4	4
Participant 5	4	2	2	4
Participant 6	5	5	5	5
TOTAL (max 30)	21	17	24	26
Median (IQR)	4 (0.75)	3 (2.5)	4 (0.75)	4 (0.75)

Table 7-6: Patient responses on feedback sheet to ‘helpfulness’ of the therapy on four major communication abilities (auditory comprehension, reading, spoken output and memory) according to the Likert Scale.

None of the patients reported using the therapy prior to this study and 83% stated they 'enjoyed' carrying out the React2© therapy regularly. In terms of ease of use, all patients stated that the touch-screen tablet was both intuitive and easy to use, and that they liked 'being able to do the therapy anywhere and at any time'. However, 67% of patients stated that they required help to complete the tasks, compared to no healthy participants requiring help. Despite this, 83% of patients stated that they enjoyed carrying out the therapy and participating in the study. 1 out of the 6 patients stated they would consider buying the therapy programme for themselves following participation in the study, with the 5 remaining patients reporting that they did not feel they required the React2© programme any longer.

Other comments relating to the React2© therapy were recorded and subjectively divided into positive and negative statements, as follows:

Positive comments	**Negative comments**
"React2 helped me so much!"	*"Very repetitive"*
"It fitted in with our home and lifestyle"	*" Too repetitive"*
"I could do them at my own pace and as much or as little each time"	*"It was repetitive, but it worked"*
"I was able to do the therapy myself and enjoyed that"	*"Boring"*
"It helped me to overcome aphasia"	*"It was boring"*
"I lacked technical training therefore it helped to have a touchscreen – a touchscreen was simpler for me to manage"	*"I found it boring"*
"Aphasia is initially scary and lonely, and this helped me"	

7.1.5. Discussion

Quality of word/picture verification task

Patient performance on the word/picture verification task showed a similar pattern overall to the pilot data presented in Chapter 6, in terms of the global score and accuracy rate per condition, but at a more impaired/ reduced level. This supports the claim (Breese and Hillis, 2004) that this task is sensitive at detecting auditory comprehension impairments. Results also show that patients with impaired auditory comprehension find the incongruent conditions more difficult than the congruent condition, while those without impaired auditory comprehension do not find any of the conditions difficult. Incongruency effects were more pronounced on the reaction time (RT) measurements with reduced RTs on the phonological and semantic conditions as compared to the correct condition, suggesting a difficulty effect on the phonological and semantic conditions (same pattern as the pilot controls in Chapter 6). However, in contradiction to Breese and Hillis' results (2004), this study suggests that this task is not sensitive at dissociating within impaired individuals when using the task as a behavioural assessment. Unless reaction times are also recorded (rarely clinically viable), these results question the clinical relevance of this task as an assessment tool.

No relationship was found between the overall Western Aphasia Battery (WAB) language score (Kertesz, 1982) and the auditory comprehension task scores/ RTs, supporting Breese and Hillis' (2004) suggestion that this task is not an overall communication assessment. The correlation found between the WAB comprehension subtest score (Kertesz, 1982) and the task was expected as both are measures of auditory comprehension. The fact that this correlation wasn't significant was also unsurprising since the WAB comprehension subtest is a generic measure of auditory comprehension (from the level of single word up to sentence level), whereas the word/picture verification task aims to assess single noun auditory comprehension. Following the same logic, it was expected that the Peabody Picture Vocabulary Test 4 (PPVT4) (Dunn and Dunn, 2007) would also show a positive correlation with the task - a result that was not seen. This could be due to the fact that although the PPVT4 tests single word auditory comprehension, it focusses on abstract nouns more than concrete ones – the opposite to both the picture/word verification task and the WAB comprehension assessment. Also of note is the fact that the PPVT4 is a multiple choice single word auditory comprehension assessment, and thus the test layout that Breese and Hillis' (2004) task was designed to improve upon.

As discussed in Chapter 6, Breese and Hillis' (2004) word/picture verification task is based on the interactive distributed model of semantic processing. Therefore, it was expected that a strong correlation would be found between scores on the task and scores on both The Pyramids and Palmtrees Test (Howard and Patterson, 1992) and The Kissing and Dancing Test (Bak and Hodges, 2003). However, no correlation was found between the task and either assessment. The lack of correlation with The Kissing and Dancing Test is most likely due to the fact that this assessment focuses on action semantics, whereas the task focusses on noun semantics. However, the lack of correlation between the task and The Pyramids and Palmtrees Test remains. It is possible that, as alluded to earlier, this lack of correlation is because the task actually has a reduced sensitivity to semantic processing than expected (Breese and Hillis, 2004).

Finally, a significant positive correlation was found between the task score and the Doors subtest of the 'Doors and People Test' (Baddeley et al., 1994). This assessment measures visual short-term memory skills, and thus a correlation was not unexpected as this skill is relied upon within task. In contrast, no correlation was found between verbal short-term working memory (Wechsler Adult Intelligence Scale-III (WAIS III): digit span subtest (Ryan and Lopez, 2001)), suggesting that the picture/word verification task relies more heavily on visual than verbal memory skills.

Lesion location/ cognitive neuro-psychological model

Following our model of single word processing (Chapter 6 – Figure 6-4), it was expected that patient performance would significantly differ according to the four conditions within the word/picture verification task. In terms of performance no strong dissociations were observed, with only a semantic incongruency effect seen ($p=0.02$, bootstrapped multi-comparisons). Phonological and semantic incongruency effects are seen on the reaction times, although there is no significant difference in these effects from the unrelated condition. Taken together, the semantic incongruency seen on the scores and reaction times suggest that the semantic condition is the most difficult part of the task for the patient population, supporting the idea that Breese and Hillis' (Breese and Hillis, 2004) word/picture verification task is a test of semantic processing.

As patients are often classified according to their anatomical damage, this study aimed to further investigate links between neuroanatomy and our model of single word processing (Chapter 6 – Figure 6-4). fMRI results from Chapter 6 suggest that the pars triangularis is involved in phonological congruency and working memory; the left angular gyrus is involved in semantic congruency; and the left precuneus is involved in semantic relatedness. Therefore, using this information, it was expected that patients who performed above chance level for the incongruent conditions would not have any lesion damage in these regions. Three patients were found matching this behavioural performance, none of which showed damage to the left angular gyrus or the left precuneus. One patient did, however, show lesion damage to the pars triangularis. Using the same rationale, it was also expected that patients who performed below chance level for the incongruent conditions on the task would have lesion damage to the left angular gyrus and/or left precuneus. Interestingly, all three patients showed damage to the pars triangularis – an area involved in phonological congruency and working memory within this task. One of these three patients (behaviourally performed below chance level) also showed lesion damage to the left precuneus and left angular gyrus. If we look at all patients showing angular gyrus lesions an additional patient shows up, who was unable to carry out the task at all (no responses). This means that the number of patients unable to perform on the task (scores below chance or no response) was n=2, both of whom had damage to the left angular gyrus. Taken together, these results show that if the angular gyrus is damaged then auditory comprehension of single nouns within this task is impaired, due to a semantic processing impairment. However, the results also show that an impairment within this task is not necessarily due to a lesion within the angular gyrus, but can also be linked to a lesion within the pars triangularis. This suggests that the connections between the angular gyrus and the pars triangularis, i.e. the arcuate fasciculus (Seghier, 2013), serve a fundamental role in semantic processing, supporting the literature discussing the involvement of the pars triangularis in the lexico-semantic system (Braun et al., 2015; Fiebach et al., 2002; Poldrack et al., 1999; Skeide et al., 2014).

Following the same logic, it was expected that patients who perform above chance level in all (correct, semantic, unrelated) apart from the phonological condition (below chance) would have lesion damage to the pars triangularis. Two patients showed this pattern of behavioural performance, one with damage to the pars triangularis, and the other with an intact pars triangularis but damage to the precuneus. As the angular gyrus and precuneus have been linked to semantic processing, it was also expected that patients who perform below chance level on the semantic and unrelated conditions, but above chance on the correct and phonological conditions, would show damage in the left angular gyrus and/or left precuneus regions. Unexpectedly, no patients showed this behavioural pattern, even though a number of patients showed damage to these areas. Therefore, this suggests that this task is not as sensitive to impairments within the auditory processing of semantic information as originally presumed (Breese and Hillis, 2004). The average performance of patients with pars triangularis damage was 39%, confirming that a lesion in the pars triangularis does not necessarily lead to difficulties with single word auditory comprehension tasks. As discussed in Chapter 6, the pars triangularis is linked to phonological working memory (Demonet et al., 1992; Liberman and Mattingly, 1985). It is therefore suggested that those showing this lesion location with an associated impairment on the task are unable to rehearse and sequence strategies when processing phonemic strings. It is also possible that those who show this lesion location with no associated impairment are able to utilise other aspects of the phonological loop to carry out the task (Muller and Knight, 2006).

Wernicke's area

During this study, it emerged that there is a broad consensus as to the function of Wernicke's area (word and sentence auditory comprehension), however its anatomical location remains controversial. Classically, it is defined as located in the posterior portion of Brodmann's area 22, but emerging evidence suggests a larger composite area involving the angular gyrus, the supramarginal gyrus and the posterior inferior, middle and superior temporal gyri (Mesulam et al., 2015).

Therefore, to investigate the function of the angular gyrus within Wernicke's area further comparative analysis was carried out looking at the task results (overall performance, accuracy rates per condition and reaction times per condition) for patients with damage (or not) to Wernicke's area, according to the classical anatomical location and according to Mesulam's (2015) composite anatomical location. Accuracy rates per condition were then investigated by robustly testing for mean differences between groups with an adjustment for multiple comparisons. Figure 7-8 shows the different performance patterns on the auditory comprehension task according to the two different definitions of Wernicke's neuroanatomical area. Although not significantly different, it is possible to see from the pattern of performances on the global scores and accuracy rates that there is an increased difference between the performance of the lesioned and non-lesioned groups when using Mesulam's (2015) classification system over the traditional one (Figure 7-8). As discussed, the results above show that the angular gyrus is necessary in order to carry out Breese and Hillis' (2004) picture/word verification task. As the angular gyrus is included in the composite classification of Wernicke's area, and not in the traditional definition, this suggests that the traditional classification of Wernicke's is not specific enough to use in anatomical classifications of aphasia. Instead, it is suggested that the anatomical classification of Wernicke's aphasia should be based on Mesulam's (Mesulam et al., 2015) definition.

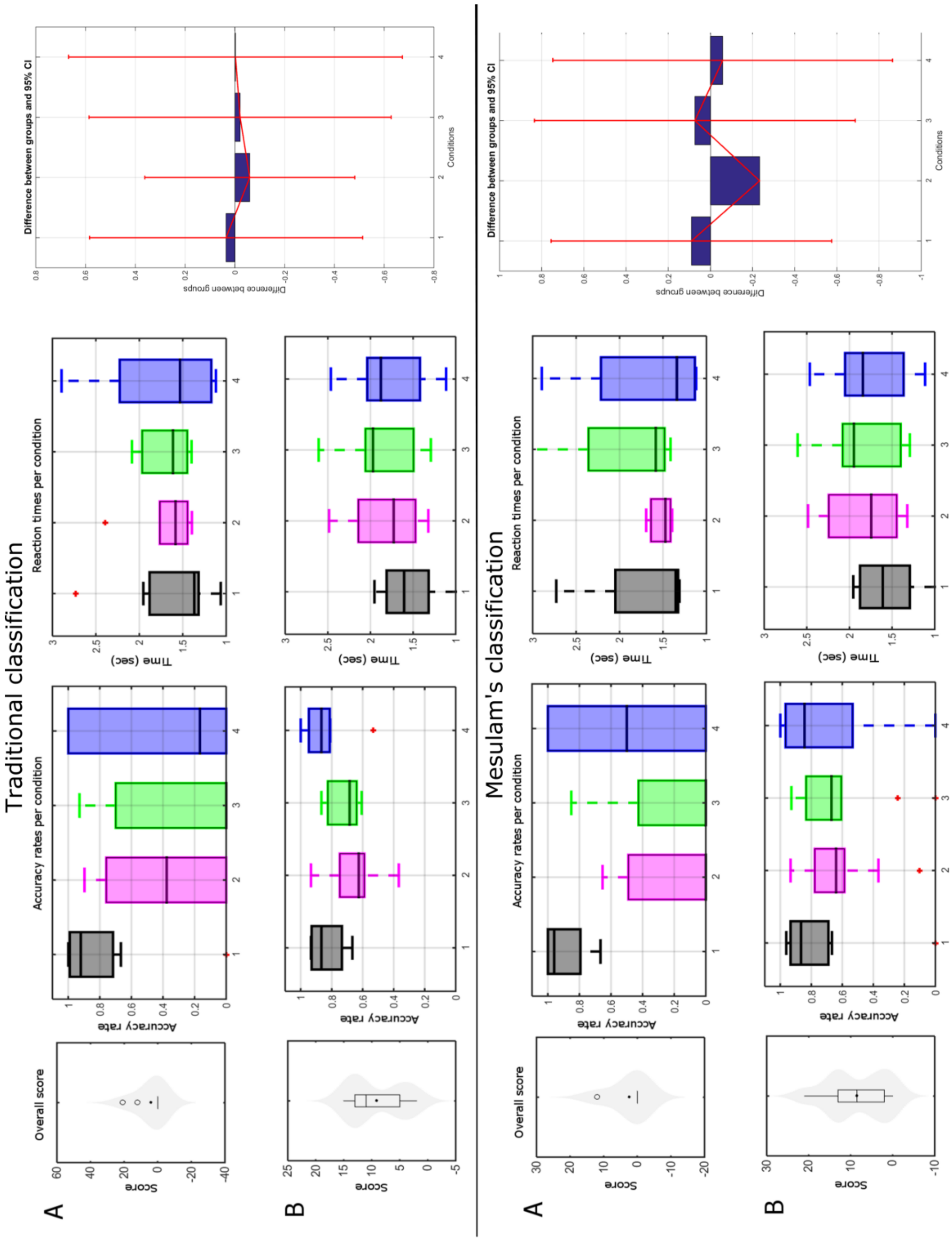

Figure 7-8: Performance differences according to classification of Wernicke's area (traditional v. composite).

A: patients with a lesion according to the traditional Wernicke's area classification (top); *B*: patients with no lesion according to Mesulam's composite classification of Wernicke's area (bottom).

Left- Right: Violin plot for global score distribution; Box plots showing accuracy rates per condition (grey: correct, pink: phonological, green: semantic, blue: unrelated); Box plots showing reaction times (sec) per condition; Graphs showing the mean difference in accuracy rates between the lesioned and non-lesioned groups on the four task conditions (1: correct, 2: phonological, 3: semantic, 4: unrelated). All comparison/ difference results were not statistically significant. Red crosses signify outliers.

Therapy

This research included a pilot study investigating the effectiveness of React2© therapy on participant outcomes by looking for correlations between therapy measures and changes in language and cognitive assessment scores. Due to a number of limitations within this study (detailed in the following section), only a few of the results reported were statistically significant, although a number of trends can be seen in our data. However, these preliminary results suggest that the greater number of exercises patients carry out on the React2© therapy, as well as greater amount of time spent carrying out the therapy, the more improvement patients will show in terms of semantic processing of single noun auditory comprehension.

Furthermore the results also suggest that patient scores on the WAB comprehension subtest showed a larger improvement when they achieved higher scores on the React2© therapy. Multiple group comparisons showed a small difference on the WAB comprehension subtest, between the healthy control and the patient group with therapy. This suggests that the easier the React2© therapy is, the more reduced the improvement in auditory comprehension skills.

For the patient group, it was expected that therapy scores would also correlate with improvements in the Peabody Picture Vocabulary Test (PPVT4) (Dunn and Dunn, 2007) as both the assessment and the therapy are focussed on auditory comprehension abilities. However, no correlation was found, suggesting that the auditory comprehension sections of the React2© therapy carried out by participants in this study were focussed at multiple levels of auditory comprehension, as compared to the specificity of the linear single noun auditory comprehension level of the PPVT4 assessment. Therefore, it is important for clinicians to tailor both the therapy and outcome assessment tools not only to a specific communication target skill, but to a specific level within this skill, and to each other.

Importantly, the feedback received from participants suggests that carrying out the React2© therapy has a positive impact on perceived outcomes in terms of specific communication skills. The importance of delivering the therapy through an 'easy-to-use' interface (touch-screen, portable, tablet computer) was emphasised by participants and should be taken into consideration when providing patients with any form of computer-assisted speech and language therapy. In feedback, 6 participants (50%) stated that the React2© therapy was either 'boring' or 'repetitive'. Clinically this is important, especially as this particular therapy programme actually contains over 8000 different exercises. Speech and Language Therapists aim to provide targeted impairment-based therapy to participants, meaning that the number of exercises suitable (specific therapy target, level of difficulty among other factors) for an individual will be greatly reduced in number, as was the case in this pilot study. Therefore, it is important that any exercises provided to an individual patient are regularly reviewed by a Speech and Language Therapist, in terms of suitability and repetition, in order to maintain high levels of motivation and interest with the patient.

Limitations/ Future work

There are some study design and patient-related limitations that arose as this research progressed, discussed below.

Recruitment to any acute stroke trial is challenging for a number of reasons, leading to a slow and reduced recruitment for this study. There were a number of reasons for this, but of the main ones identified was the strict inclusion and exclusion criteria. Although the criteria were reviewed part-way through recruitment, they were not changed so that the main aims and purposes of the study could be accomplished. If the criteria had been altered it may have decreased the likelihood of producing results that are scientifically reliable and reproducible. As the patient group involved in this study are particularly vulnerable, due to communication impairments, it was important to guard against any exploitation and keep the risk of harm to absolute minimum. Timing was also a factor that influenced patient recruitment – as this study aimed to look at auditory comprehension and therapy effectiveness at the sub-acute phase, this meant that recruitment was carried out when patients were at the acute phase following stroke. This left a very limited time window for patient recruitment and so limiting eligibility. As the patients were actively sick at this time point it was often a stressful time for the patient and family members, thus a number of potential participants declined recruitment due to being overwhelmed by the situation. The nature of this study meant that all patients targeted for recruitment to the study had an auditory comprehension impairment of some degree. Due to ethical considerations protocol required that all participants were able to fully consent themselves, again reducing the number of eligible patients dramatically. Importantly, this did not impact upon trial generalisability for the treatment section of this study, as participants were required to have a minimum level of comprehension, compliance and cognition to participate in assessments and therapy. Interestingly, patient group A (i.e. those randomised to the therapy group) had a full rate of follow-up (100%), whereas patient group B (i.e. those randomised to the no-therapy group) showed a significantly reduced rate of follow-up compliance (56%). Although there were 3 participants that had unavoidable medical reasons to withdraw from the study, 2 participants stated that as they weren't getting any extra therapy they felt the study was too time-consuming (*note*: 1 participant from patient group B did not complete the feedback). Therefore, in future it would be useful to reduce the amount of contact time with participants involved in such a study by reducing the number of assessments carried out at each time point. It would also be useful, as well as recording the number of patients opting into the study, to document the number of patients approached, and number of patients declined/reasons for opting out.

In terms of study design, the assessment battery used was intentionally constrained to investigate auditory comprehension at the specific levels being researched within this thesis (phoneme and single-word level) and did not cover higher levels of auditory comprehension, for example syntactical structure. In future, it would be interesting to see if any changes in phoneme/ single-word level auditory comprehension could be generalised to higher levels, for example by using sections of the CAT

(Comprehensive Aphasia Test - (Swinburn et al., 2004)) or the PALPA (Psycholinguistic Assessments of Language Processing in Aphasia - (Kay et al., 1992)). Another limitation of the study design is that it was difficult to determine whether any results from the therapy group were due to spontaneous recovery or due to the therapy. This is partly due to the small sample sizes recruited for the study and partly due to the timing of delivery of therapy (sub-acute time point – when the majority of spontaneous recovery occurs (Alexander, 1994; Mazzocchi and Vignolo, 1979)). In future it would be interesting to carry out a comparative study with patients at the chronic phase of recovery. The patient group included was deliberately homogenous, however it would be interesting to carry out a similar study with a more heterogeneous patient cohort.

There are a number of therapy outcomes that were beyond the scope of this study, but would be interesting to look at in future investigations. These include the emotional impact of independent home-based therapy as opposed to conventional face-to-face therapy, the emotional impact of the provision of extra therapy resources, and the cost effectiveness of computer therapy (Kearns, 1993; Mortley et al., 2004; Palmer et al., 2012). There was no functional assessment of communication abilities included in the battery of language assessments on the patient population. It would be interesting to include an assessment such as the CADL-2 (Communicative Abilities in Daily Living-2 (Holland et al., 1999)) in order to investigate the impact of any therapy on participants' lives that may not be obvious through conventional language assessments (Kearns, 1993).

Conclusions

This study shows that the word/picture verification task is good at detecting auditory comprehension impairments, but that it lacks sensitivity in terms of differentiation amongst these impairments. The task also relies upon visual working memory, and thus this must be accounted for in any task results.

Lesions within the pars triangularis do not necessarily cause difficulties with phonological processing skills of single words. However, damage to this area can cause impairments in phonological processing, through the suggested route of damage to the phonological working memory. Lesions within the pars triangularis do not necessarily cause difficulties with processing of single words, but damage to this area can cause impairments in semantic processing. Lesions within the left angular gyrus do necessarily cause impairments in single word comprehension. Therefore, it is suggested that connections between the angular gyrus and the pars triangularis serve a fundamental role in semantic processing.

Finally, preliminary results show that utilising React2© computer-assisted therapy independently for auditory comprehension impairments may be effective. Results suggest that the more therapy patients carry out the more improvement is seen in auditory comprehension skills in terms of semantic processing of single nouns. Importantly, due to a number of limitations these results need to be reproduced on a larger scale in order to be able to generalise the outcomes.

8. Chapter 8 – Discussion

8.1. Conclusions

Speech perception cannot be solely understood in terms of acoustic/ gestural events and the perception of phonemes, as many other sources and levels of information come into play. This implies that auditory comprehension, from the level of the voice up to single words, is a multifaceted process.

In the first study presented (Chapter 3), an interaction between the perceptual processes used to code voice information and those used to code phonemic information in the speech signal was shown. Two 2 AFC identification tasks were used to investigate the categorical perception of voice gender (male-female) and speech sounds (/pæ/-/tæ/) in the healthy population. In contradistinction with the classic speech perception hypothesis of speaker normalisation (suggesting a dissociation between voice/ speech processing as voice specific information is stripped away first to access phonemic information) this study showed that phonemes are categorised faster than voice, supported by the fact that allophonic variation does not impair comprehension (Mitterer et al., 2008). The reaction time differences can be explained by a theory of long-term memory interactions derived from very early stages of selection/ enhancement of important features from aurally presented information. It is suggested that these early memory interactions create expectations that enhance both the formants of the consonant in the phoneme task, and the f0 and formants of the vowel in the gender task. This interpretation is supported by the reverse correlation analysis that suggests that gender and phoneme are processed on the basis of different perceptual representations. It is proposed that phoneme categorisation requires fine temporal analysis, relying more on the left hemisphere (Cohen, 2012; Poeppel and Hickok, 2004; Zatorre et al., 2002), specifically the mid-STS (phoneme categorisation and voice recognition). In contrast, gender categorisation requires a finer spectral analysis, relying more on the right hemisphere (Charest et al., 2012; Zatorre et al., 2002), specifically the anterior STS which analyses specific gender features after speech information has been processing (functional integration). In the follow-up study (Chapter 4), a stroke patient cohort partly confirmed this hypothesis. Lesion-symptom mapping showed that lesions of the right frontal cortex (likely the ventral IFG) lead to voice perception deficits, whereas left fronto-temporal lesions showed a tendency for associated performances (both voice and speech perception impaired or both not). Together these results support the hypothesis of bilateral processing of voice information (Belin et al., 2000; Belin et al., 2002; Fecteau et al., 2004; Fecteau et al., 2005), with an important role of the right frontal cortex in carrying out voice categorisation, and common mechanisms within the left hemisphere for carrying out talker and speech information. Currently, it is proposed that there are cognitive voice processing models (Belin et al., 2004) and separate cognitive models of speech and single word processing (Whitworth et al., 2014), but only Specht's (2014) model associates both processes. This is most likely due to the fact that the early stages of single word processing (i.e. auditory phonological analysis) are still highly debated in terms of the specificity versus generalisation of pre-lexical processing. The data presented within this thesis suggests that voice processing is an integral part of single word processing. As results from Chapters 3 and 4 show that gender (voice) and phoneme perception are processed on the basis of different perceptual representations, it is proposed that when an individual hears a word, general auditory acoustic analysis occurs first, prior to any specific analysis (i.e. phoneme or voice level). Furthermore, results from these studies 1. show an extended reaction time for phoneme perception as compared to voice perception and 2. suggest a possible association of phoneme and voice categorisation. Therefore, it is proposed that following this general processing level, acoustic phonological and acoustic voice analysis occur in parallel, with the voice analysis module interacting (bi-directionally) with the phonological input lexicon (Figure 8-1). The possible association/interaction between voice and phoneme perception also suggests that commencing Speech and Language Therapy at a low level of acoustic processing/ voice perception may be an appropriate method in the treatment for patients showing phoneme perception impairments.

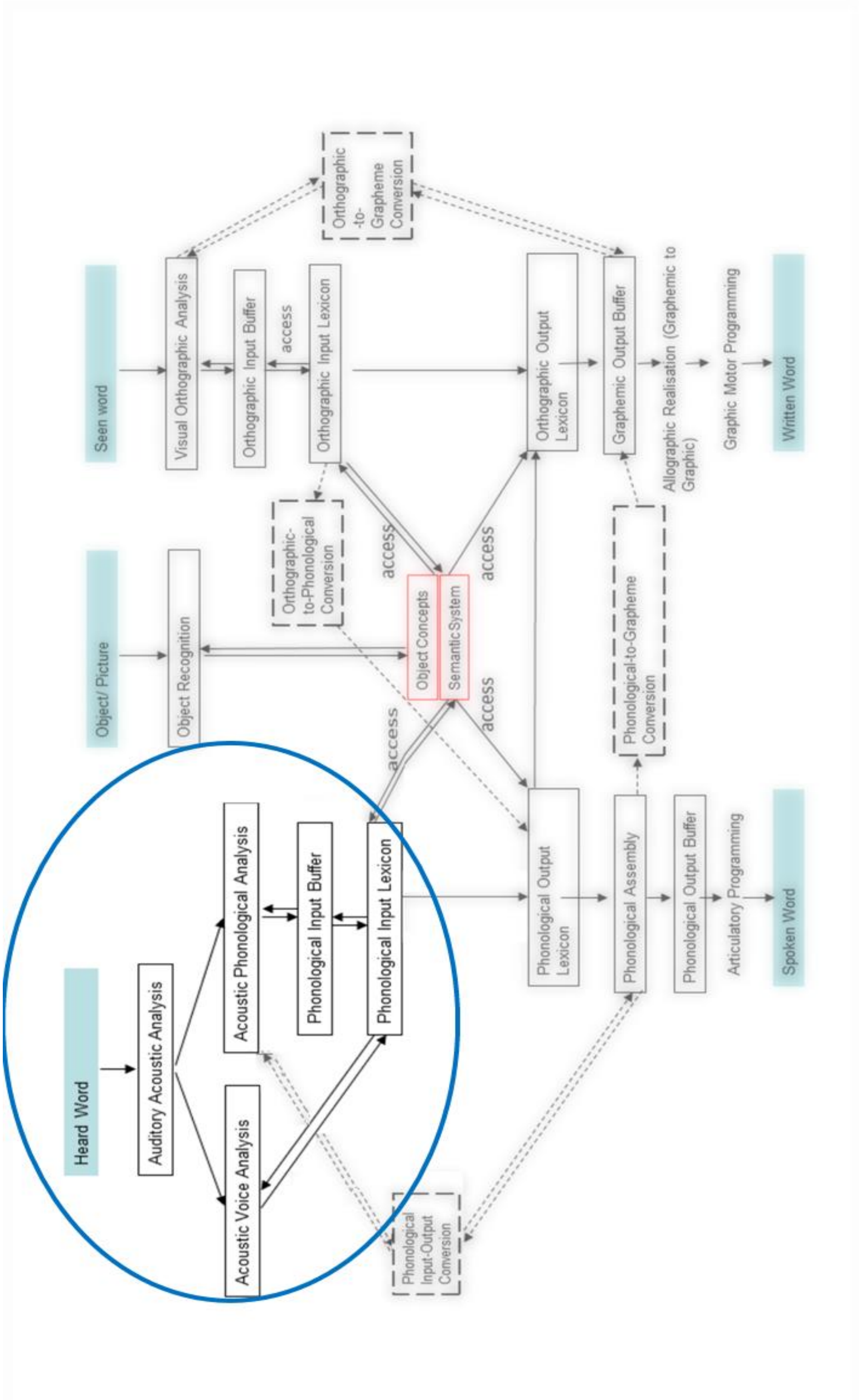

Figure 8-1: Suggested addition of voice processing into the adapted and extended cognitive neuropsychological model (Figure 2-8).

It is proposed that general auditory acoustic analysis is added prior to the acoustic phonological analysis. Acoustic voice analysis then occurs in parallel with acoustic phonological analysis, and interacts (bi-directionally) with the phonological input lexicon.

In depth behavioural analysis of a longitudinal case study (Chapter 5) of crossed aphasia showed good concordance with the adapted and extended version of the cognitive neuropsychological model of single word processing (Chapter 2, Figure 2-8), supporting the utility of this model for research and clinical purposes. While results suggest that crossed aphasia can be explained by developmental disorders causing partial right lateralization shift of language processes, they also revealed that the shift did not affect voice lateralization, revealing at least partial segregation of these processes. Specht's (2014) model of triple-stream auditory comprehension processing is one of the only neuroanatomical models currently available that includes both voice and speech processing. This model suggests a right lateralised ventral stream of processing focussing on voice processing, a left lateralised ventral stream focussing on lexical processing, semantics and syntax, as well as a left lateralised dorsal stream focussing on auditory-motor integration. The single case of crossed aphasia presented supports Specht's (2014) proposed right hemisphere specialization for voices – while the other results presented suggest a more associated nature of voice/speech in the left hemisphere.

Moving up to single word comprehension, functional MRI results of a word-picture verification task suggest that the pars triangularis is involved in phonological processing (most likely phonological working memory), while left angular gyrus was involved in semantic incongruency decisions, and the left precuneus was linked to semantic relatedness. The follow-up study in stroke patients further showed that connections between the angular gyrus and the pars triangularis are fundamental in semantic processing of single words, at least within this word-picture verification task. According to Specht's (2014) cognitive neuroanatomical model, the left lateralised ventral stream is involved in sub-lexical, lexical, semantic and syntactical processing. The model proposes that this stream connects the angular gyrus with the inferior frontal gyrus via the middle and inferior temporal gyrus/ superior temporal sulcus, and the superior temporal gyrus/temporal pole/anterior temporal lobe. Our results support this idea, additionally suggesting an association with a visuo-spatial area (precuneus). Despite this refinement in anatomo-functional model, the data presented in Chapter 7 shows that is it is uncommon for patients to fit neatly into anatomical and behavioural models, however these remain helpful ways of classifying patients for both therapy and research purposes.

8.2. Limitations

8.2.1. Participants

There was a potential selection bias of the healthy controls participating within the studies presented, as recruitment was mainly from within the University of Edinburgh. Even still, there remained a variety of age, education and sex within all healthy control groups as the student population from within the University was solely used. Within the patient populations recruited for these studies the participant sampling bias was more reduced in terms of age, education and sex. However, recruitment of stroke patients causes its own difficulties, especially when recruiting patients with limited communication abilities (aphasia). The speed of recruitment of patient participants limited the number of participants that were able to take part in the studies overall. This was due to a number of factors including inclusion and exclusion criteria (necessary in order to retain high scientific value of the studies), financial restraints (reduced transport options available for participants), patient or carer resistance (especially at the acute phase) and reduced patient understanding (aphasia).

8.2.2. Imaging

Imaging within the study protocols occasionally dissuaded patients from participating, especially if the patients were at the acute time point. Patients often commented that MRI scanning was uncomfortable. NHS clinical structural scans collected and analysed within these studies were a combination of CT and MRI, leading to a lack of consistency when carrying out any lesion mapping.

8.2.3. Therapy

For all studies that included patients the frequency, intensity and duration of Speech and Language Therapy, as well as the specific assessments received, was not data that was readily available in an accurate format. Within the React2© therapy, importantly ceiling level was reached quickly by all healthy controls (as anticipated) but also by the majority of participants, suggesting that the highest levels within this therapy programme are not complex enough for a number of patients with post-stroke auditory comprehension impairments.

8.3. Future work

This thesis expands on the current knowledge regarding auditory comprehension at the levels of the voice, phoneme and single word processing. However, this remains a complex, highly debated and extensive field of work which requires further research in all areas.

Further work investigating the proposed addition of the voice perception module into the current cognitive neuropsychological model of single word processing is possible. Clinically, such an addition would hold implications for the assessment and therapy provided to those with auditory comprehension impairments. Furthermore, the specific neuroanatomy underlying voice perception remains debated and requires further investigation.

In terms of the proposed combination of the cognitive neuropsychological and neuroanatomical models of single word auditory comprehension, these theories require further investigation in larger scale studies, with more consistent imaging techniques (i.e. no inclusion of CT scans).

The pilot study included in the final chapter (7) of this body of work showed promising results for the use of React2© therapy as an independent computer-assisted speech and language therapy tool for people with auditory comprehension impairments associated with post-stroke aphasia. A larger scale study investigating the effectiveness of this therapy would be appropriate. However, prior to carrying out such a study it is important to address the limitations that arose within the pilot study. Patient recruitment issues would need to be addressed with larger numbers of researchers working to recruit within such a study. As recruitment at the acute stage of stroke recovery presents its own set of complications it would be worth considering recruiting at the sub-acute/chronic phase of recovery. It would also be important to address the ceiling limitations of the therapy, by working with the creators (http://www.react2.com/) to provide additional, higher levels within those models addressing auditory comprehension impairments.

Reference List

MATLAB version R2014b. 2014. Natick, Massachusetts, The MathWorks Inc.

Aaltonen,O., Tuomainen,J., Laine,M., and Niemi,P. (1993). Cortical differences in tonal versus vowel processing as revealed by an ERP component called mismatch negativity (MMN). Brain and Language *44*, 139-152.

Acheson,D., and Hagoort,P. (2013). Stimulating the brain's language network: syntactic ambiguity resolution after TMS to the inferior frontal gyrus and middle temporal gyrus. Journal of Cognitive Neuroscience *25*, 1664-1677.

Ahrens,M., Hasan,B., Giordano,B., and Belin,P. (2014). Gender differences in the temporal voice areas. Frontiers in Neuroscience *8*, 228.

Alajouanine,T., Ombredane,A., and Durand,M. (1939). Le syndrome de desintegration phonetique dans l'aphasie (Paris: Masson).

Alexander,M. (1994). Stroke Rehabilitation Outcome: A Potential Use of Predictive Variables to Establish Levels of Care. Stroke *25*, 128-134.

Alexander,M., and Annett,M. (1996). Crossed aphasia and related anomalies of cerebral organization: case reports and a genetic hypothesis. Brain and Language *55*, 213-239.

Alho,K., Vorobyev,V., Medvedev,S., Pakhomov,S., Starchenko,M., Tervaniemi,M., and Naatanen,R. (2006). Selective attention to human voice enhances brain activity bilaterally in the superior temporal sulcus. Brain research *1075*, 142-150.

Alles,R., Bamiou,D., Batchelor,L., Campbell,N., Canning,D., Grant,P., Luxon,L., Moore,D., Murray,P., Nairn,S., Rosen,S., Sirimanna,T., Treharne,D., and Wakeham,K. Auditory Processing Disorder (APD): Position Statement. 2011. The British Society of Audiology.

Altmann,G. (1995). Cognitive models of speech processing: Psycholinguistic and computational perspectives (Massachusetts: MIT Press).

American Psychiatric Association (2013). Diagnostic and statistical manual of mental disorders (Washington: American Psychiatric Publishing).

Andics,A., Gacsi,M., Farago,T., Kis,A., and Miklosi,A. (2014). Voice-sensitive regions in the dog and human brain are revealed by comparative fMRI. Current Biology *24*, 574-578.

Annett,M. (1985). Left, right, hand and brain: The right shift theory (Virginia: Erlbaum).

April,R., and Han,M. (1980). Crossed aphasia in a right-handed bilingual Chinese man: A second case. Arch Neurol *37*, 342-346.

Ardila,A., and Rosselli,M. (2002). Acalculia and dyscalculia. Neuropsychology Review *12*, 179-231.

Assal,G., Buttet,J., and Jolivet,R. (1981a). Dissociations in aphasia: a case report. Brain and Language *13*, 223-240.

Assal,G., and Hadj-Djilani,M. (1976). Une nouvelle observation d'alexie pure sans hemianopsie. Cortex *12*, 169-174.

Assal,G., Perentes,E., and Deruaz,J. (1981b). Crossed aphasia in a right-handed patient: Postmortem findings. Arch Neurol *38*, 455.

Awad,M., Warren,J., Scott,S., Turkheimer,F., and Wise,R. (2007). A common system for the comprehension and production of narrative speech. The Journal of Neuroscience *27*, 11455-11464.

Bachorowski,J., and Owren,M. (1999). Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. The Journal of the Acoustical Society of America *106*, 1054-1063.

Baddeley,A., Emslie,H., and Nimmo-Smith,I. (1994). The Doors and People Test (Bury St. Edmunds: Thames Valley Test Company).

Bak,T., and Hodges,J. (2003). Kissing and Dancing - a test to distinguish the lexical and conceptual contributions to noun/verb and action/object dissociation. Preliminary results in patients with frontotemporal dementia. Journal of Neurolinguistics *16*, 169-181.

Bakar,M., Kirshner,H., and Wertz,R. (1996). Crossed aphasia: functional brain imaging with PET or SPECT. Arch Neurol *53*, 1026-1032.

Bakheit,A., Shaw,S., Barrett,L., Wood,J., Carrington,S., Griffiths,S., Searle,K., and Koutsi,F. (2007). A prospective, randomized, parallel group, controlled study of the effect of intensity of speech and language therapy on early recovery from poststroke aphasia. Clinical Rehabilitation *21*, 885-894.

Bamiou,D., Musiek,F., Stow,I., Stevens,J., Cipolotti,L., Brown,M., and Luxon,L. (2006). Auditory temporal processing deficits in patients with insular stroke. Neurology *67*, 614-619.

Barnett,V., and Lewis,T. (1994). Outliers in statistical data (New York: Wiley).

Bartha,L., Marien,P., Poewe,W., and Benke,T. (2004). Linguistic and neuropsychological deficits in crossed conduction aphasia. Report of three cases. Brain and Language *88*, 83-95.

Basso,A., Capitani,E., and Laiacona,M. (1995). Primary Progressive Aphasia and Semantic Dementia. Progressive language impairment without dementia: a case with isolated category specific semantic defect. Neurocase *1*, 39-54.

Basso,A., Lecours,A., Moraschini,S., and Vanier,M. (1985). Anatomoclinical correlations of the aphasias as defined through computerized tomography: Exceptions. Brain and Language *26*, 201-229.

Bates,E., Wilson,S., Saygin,A., Dick,F., Sereno,M., Knight,R., and Dronkers,N. (2003). Voxel-based lesion-symptom mapping. Nat Neurosci *6*, 448-450.

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

Belin,P., and Grosbras,M. (2010). Before speech: Cerebral voice processing in infants. Neuron *65*, 733-735.

Belin,P., and Zatorre,R. (2003). Adaptation to speaker's voice in right anterior temporal lobe. NeuroReport *14*.

Belin,P., Zatorre,R., Lafaille,P., Ahad,P., and Pike,B. (2000). Voice-selective areas in human auditory cortex. Nature *403*, 309-312.

Belin,P., Zatorre,R.J., and Ahad,P. (2002). Human temporal-lobe response to vocal sounds. Brain Res. Cogn Brain Res. *13*, 17-26.

Benson,D. (1979). Aphasia, alexia, and agraphia (New York: Churchill Livingstone).

Benson,R., Richardson,M., Whalen,D., and Lai,S. (2006). Phonetic processing areas revealed by sinewave speech and acoustically similar non-speech. NeuroImage *31*, 342-353.

Benson,R., Whalen,D., Richardson,M., Swainson,B., Clark,V., Lai,S., and Liberman,A. (2001). Parametrically dissociating speech and nonspeech perception in the brain using fMRI. Brain and Language *78*, 364-396.

Berninger,V., Abbott,R., Thomson,J., and Raskind,W. (2001). Language phenotype for reading and writing disability: A family approach. Scientific Studies of Reading *5*, 59-106.

Berthier,M., Davila,G., Garcia-Casares,N., Green,C., Juarez,R., Ruiz-Cruces,R., Pablo,L., and Barbancho,M. (2011). Atypical conduction aphasia and the right hemisphere: cross-hemispheric plasticity of phonology in a developmentally dyslexic and dysgraphic patient with early left frontal damage. Neurocase *17*, 93-111.

Bestelmeyer,P., Belin,P., and Grosbras,M. (2011). Right temporal TMS impairs voice detection. Current Biology *21*, R838-R839.

Bhatnagar,S., Buckingham,H., Puglisi-Creegan,S., and Hacein-Bey,L. (2011). Crossed aphasia in a patient with congenital lesion in the right hemisphere. Aphasiology *25*, 27-42.

Bhatnagar,S., Imes,S., Buckingham,H., and Puglishi-Creegan,T. (2006). Anomalous crossed aphasia in a patient with congenital lesion in the right hemisphere. Brain and Language *99*, 61-62.

Bhogal,S., Teasell,R., and Speechley,M. (2003). Intensity of Aphasia Therapy, Impact on recovery. Stroke *34*, 987-993.

Binder,J., Frost,J., Hammeke,T., Bellgowan,P., Springer,J., Kaufman,J., and Possing,E. (2000). Human Temporal Lobe Activation by Speech and Nonspeech Sounds. Cerebral Cortex *10*, 512-528.

Binder,J., Frost,J., Hammeke,T., Cox,R., Rao,S., and Prieto,T. (1997). Human brain language areas identified by functional magnetic resonance imaging. The Journal of Neuroscience *17*, 353-362.

Bishop,C. (1995). Neural networks for pattern recognition (Oxford: Clarendon Press).

Blumstein,S., Baker,E., and Goodglass,H. (1977). Phonological factors in auditory comprehension in aphasia. Neuropsychologia *15*, 19-30.

BNC Consortium. The British National Corpus: Version 3. Oxford University Computing Services. [BNC XML Edition]. 2007.

Boer,E. (1976). On the "Residue" and Auditory Pitch Perception. In Handbook of Sensory Physiology: Clinical and Special Topics, (Berlin: Springer), pp. 479-583.

Boersma,P., and Weenink,D. Praat: doing phonetics by computer. [5.1.05]. 2009.

Boller,F., Youngjai,K., and Mack,J. (1977). Auditory Comprehension in Aphasia. In Studies in Neurolinguistics: Perspectives in Neurolinguistics and Psycholinguistics, H. Whitaker, and H.A. Whitaker, eds. Academic Press), pp. 1-63.

Bonilha,L., and Fridriksson,J. (2009). Subcortical damage and white matter disconnection associated with non-fluent speech. Brain *132*, e108.

Bozeat,S., Lambon Ralph,M., Patterson,K., Garrard,P., and Hodges,J. (2000). Non-verbal semantic impairment in semantic dementia. Neuropsychologia *38*, 1207-1215.

Bradlow,A., Kraus,N., Nicol,T., McGee,T., Cunningham,J., Zecker,S., and Carrell,T. (1999). Effects of lengthened formant transition duration on discrimination and neural representation of synthetic CV syllables by normal and learning-disabled children. The Journal of the Acoustical Society of America *106*, 2086-2096.

Brady,M., Kelly,H., and Enderby,P. Cochrane Review: Speech and language therapy for aphasia following stroke. 2012. John Wiley and Sons Ltd.

Brainard,D. (1997). The psychophysics toolbox. Spatial vision *10*, 433-436.

Bramwell,B. (1899). On "Crossed" aphasia and the factors which go to determine whether the "leading" or "driving" speech-centre shall be located in the left or in the right hemisphere of the brain. The Lancet *153*, 1473-1479.

Braun,M., Hutzler,F., Munte,T., Rotte,M., Dambacher,M., Richlan,F., and Jacobs,A. (2015). The neural bases of the pseudohomophone effect: Phonological constraints on lexico-semantic access in reading. Neuroscience *295*, 151-163.

Breese,E., and Hillis,A. (2004). Auditory comprehension: Is multiple choice really good enough? Brain and Language *89*, 3-8.

Broca,P. (1863). Localization of cerebral functions. Location of articulate language. Bulletin of the Society of Anthropology *4*, 200-203.

Brodeur,M., Dionne-Dostie,E., Montreuil,T., and Lepage,M. (2010). The Bank of Standardized Stimuli (BOSS): A New Set of 480 Normative Photos of Objects to Be Used as Visual Stimuli in Cognitive Research. PLoS ONE *5*, e10773.

Brown,J., and Wilson,F. (1973). Crossed aphasia in a dextral: A case report. Neurology *23*, 907-911.

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

Brumfitt,S., and Sheeran,P. (1999). Visual Analogue Self-Esteem Scale (VASES) (Oxon: Winslow Press Ltd.).

Buccino,G., Binkofski,F., Fink,G., Fadiga,L., Fogassi,L., Gallese,V., Seitz,R., Zilles,K., Rizzolatti,G., and Freund,H.-J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. European journal of neuroscience *13*, 400-404.

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

Burton,A., Bruce,V., and Johnston,R. (1990). Understanding face recognition with an interactive activation model. British Journal of Psychology *81*, 361-380.

Bybee,J., and Hopper,P. (2001). Frequency and the Emergence of Linguistic Structure (Philadelphia: John Benjamins Publishing Company).

Calhoun,V., Stevens,M., Pearlson,G., and Kiehl,K. (2004). fMRI analysis with the general linear model: removal of latency-induced amplitude bias by incorporation of hemodynamic derivative terms. NeuroImage *22*, 252-257.

Callejas,A., Shulman,G., and Corbetta,M. (2014). Dorsal and ventral attention systems underlie social and symbolic cueing. Journal of Cognitive Neuroscience *26*, 63-80.

Cao,F., Bitan,T., Chou,T., Burman,D., and Booth,J. (2006). Deficient orthographic and phonological representations in children with dyslexia revealed by brain activation patterns. Journal of Child Psychology and Psychiatry *47*, 1041-1050.

Cappa,S., Perani,D., Bressi,S., Paulesu,E., Franceschi,M., and Fazio,F. (1993). Crossed aphasia: a PET follow up study of two cases. Journal of Neurology, Neurosurgery & Psychiatry *56*, 665-671.

Caramazza,A. (1997). How many levels of processing are there in lexical access? Cognitive neuropsychology *14*, 177-208.

Carreiras,M., Carr,L., Barber,H., and Hernandez,A. (2010). Where syntax meets math: Right intraparietal sulcus activation in response to grammatical number agreement violations. NeuroImage *49*, 1741-1749.

Carrow-Woolfolk,E. (2014). Test for auditory comprehension of language (London: Pro-Ed).

Cavanna,A., and Trimble,M. (2006). The precuneus: a review of its functional anatomy and behavioural correlates. Brain *129*, 564-583.

Celsis,P., Boulanouar,K., Doyon,B., Ranjeva,J., Berry,I., Nespoulous,J., and Chollet,F. (1999). Differential fMRI responses in the left posterior superior temporal gyrus and left supramarginal gyrus to habituation and change detection in syllables and tones. NeuroImage *9*, 135-144.

Charest,I. Hierarchical organisation of voice and voice gender perception. 2009. Department of Psychology, University of Glasgow.

Charest,I., Pernet,C., Latinus,M., Crabbe,F., and Belin,P. (2012). Cerebral processing of voice gender studied using a continuous carryover fMRI design. Cerebral Cortex bhs090.

Cherney,L. (2010). Oral reading for language in aphasia (ORLA): Evaluating the efficacy of computer-delivered therapy in chronic nonfluent aphasia. Topics in stroke rehabilitation *17*, 423-431.

Chittka,L., and Brockmann,A. (2005). Perception Space - The Final Frontier. PLoS Biology *3*, e137.

Christensen,A. (1997). Communication in relation to self-esteem. Aphasiology *11*, 727-734.

Code,C., and Herrmann,M. (2003). The relevance of emotional and psychosocial factors in aphasia to rehabilitation. Neuropsychological Rehabilitation *13*, 109-132.

Cohen,H. (2012). The perceptual representations of speech in the cerebral hemispheres. The Handbook of the Neuropsychology of Language *2*, 20.

Cohen,L., Dehaene,S., Naccache,L., Lehericy,S., Dehaene-Lambertz,G., Henaff,M., and Michel,F. (2000a). The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. Brain *123*, 291-307.

Cohen,L., Dehaene,S., Naccache,L., Lehricy,S., Dehaene-Lambertz,G., Henaff,M., and Michel,F. (2000b). The visual word form area. Brain *123*, 291-307.

Cohen,L., Grony,C., Hermine,O., Gray,F., and Degos,J. (1993). Crossed aphasia with visceral situs inversus. Ann Neurol. *33*, 215-218.

Coleman,J. Discovering the acoustic correlates of phonological contrasts. ITRW on Temporal Integration in the Perception of Speech. 2002. 4-8-2002.

Coppens,P., and Robey,R. (1992). Crossed aphasia: new perspectives. Aphasiology *6*, 585-596.

Corbetta,M., and Shulman,G. (2011). Spatial neglect and attention networks. Annu. Rev. Neurosci. *34*, 569.

Coughlan,A., Oddy,M., and Crawford,A. (2007). BIRT memory and information processing battery (BMIPB). Psychology Special Interest Group for the Elderly Newsletter 29.

Crinion,J., Lambon-Ralph,M., Warburton,E., Howard,D., and Wise,R. (2003). Temporal lobe regions engaged during normal speech comprehension. Brain *126*, 1193-1201.

Csepe,V., Osman-Sagi,J., Molnar,M., and Gosy,M. (2001). Impaired speech perception in aphasic patients: event-related potential and neuropsychological assessment. Neuropsychologia *39*, 1194-1208.

Cutting,J., and Rosner,B. (1974). Categories and boundaries in speech and music. Perception & Psychophysics *16*, 564-570.

Damasio,H., and Damasio,A. (1980). The anatomical basis of conduction aphasia. Brain *103*, 337-350.

Danek,A., Gade,M., Lunardelli,A., and Rumiati,R. (2013). Tomato and Tuna: A Test for Language-Free Assessment of Action Understanding. Cognitive and Behavioral Neurology *26*, 208-217.

De Witte,L., Verhoeven,J., Engelborghs,S., De Deyn,P., and Marien,P. (2008). Crossed aphasia and visuo-spatial neglect following a right thalamic stroke: a case study and review of the literature. Behavioural neurology *19*, 177-194.

Dehaene,S., Piazza,M., Pinel,P., and Cohen,L. (2003). Three parietal circuits for number processing. Cognitive neuropsychology *20*, 487-506.

Dehaene,S., and Cohen,L. (2011). The unique role of the visual word form area in reading. Trends in Cognitive Sciences *15*, 254-262.

Dehaene-Lambertz,G., Pallier,C., Serniclaes,W., Sprenger-Charolles,L., Jobert,A., and Dehaene,S. (2005). Neural correlates of switching from auditory to speech perception. NeuroImage *24*, 21-33.

Dell,G. (1986). A spreading-activation theory of retrieval in sentence production. Psychological Review *93*, 283.

Demonet,J., Chollet,F., Ramsay,S., Cardebat,D., Nespoulous,J., Wise,R., Rascol,A., and Frackowiak,R. (1992). The anatomy of phonological and semantic processing in normal subjects. Brain *115*, 1753-1768.

Demonet,J., Price,C., Wise,R., and Frackowiak,R. (1994). Differential activation of right and left posterior sylvian regions by semantic and phonological tasks: a positron-emission tomography study in normal human subjects. Neuroscience letters *182*, 25-28.

Denes,G., and Caviezel,F. (1981). Dichotic Listening in Crossed Aphasia: 'Paradoxical' Ipsilateral Suppression. Arch Neurol *38*, 182-185.

Devlin,J., and Watkins,K. (2007). Stimulating language: insights from TMS. Brain *130*, 610-622.

di Pellegrino,G., Fadiga,L., Fogassi,L., Gallese,V., and Rizzolatti,G. (1992). Understanding motor events: a neurophysiological study. Experimental Brain Research *91*, 176-180.

Diehl,R. (1987). Auditory constraints on speech perception. In The Psychophysics of Speech Perception, (Netherlands: Springer), pp. 210-219.

Diehl,R., and Kluender,K. (1989). On the objects of speech perception. Ecological Psychology *1*, 121-144.

Diehl,R., Lotto,A., and Holt,L. (2004). Speech Perception. Annu. Rev. Psychol. *55*, 149-179.

Docherty,G., and Foulkes,P. (2004). Speaker, Speech, and Knowledge of Sounds. In Phonological Knowledge: Conceptual and Empirical Issues, N. Burton-Roberts, P. Carr, and G. Docherty, eds. (Oxford: Oxford University Press), pp. 105-130.

Dronkers,N., Plaisant,O., Iba-Zizen,M., and Cabanis,E. (2007). Paul Broca's historic cases: high resolution MR imaging of the brains of Leborgne and Lelong. Brain *130*, 1432-1441.

Dunn,L., and Dunn,D. (2007). Peabody Picture Vocabulary Test – Fourth Edition (PPVT-4) (USA: AGS Publishing).

Eickhoff,S., Stephan,K., Mohlberg,H., Grefkes,C., Fink,G., Amunts,K., and Zilles,K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. NeuroImage *25*, 1325-1335.

Eimas,P., Siqueland,E., Jusczyk,P., and Vigorito,J. (1971). Speech perception in infants. Science *171*, 303-306.

Eling,P., and Whitaker,H. (2009). History of aphasia: From brain to language. Handbook of Clinical Neurology: History of Neurology *95*, 571-582.

Ellis,A., and Young,A. (2004). Human Cognitive Neuropsychology: A Textbook with Readings (Hove: Psychology Press).

Ellis,A., Young,A., and Anderson,C. (1988). Modes of word recognition in the left and right cerebral hemispheres. Brain and Language *35*, 254-273.

Ellis,H., Jones,D., and Mosdell,N. (1997). Intra- and inter- modal repetition priming of familiar faces and voices. British Journal of Psychology *88*, 143-156.

Ellis-Hill,C., and Horn,S. (2000). Change in identity and self-concept: a new theoretical approach to recovery following a stroke. Clinical Rehabilitation *14*, 279-287.

Enderby,P., Pickstone,C., John,A., Fryer,K., Cantrell,A., and Papaioannou,D. Resource Manual for Commissioning and Planning Services for SLCN. 2009. Royal College of Speech and Language therapists.

Ethofer,T., Bretscher,J., Gschwind,M., Kreifelts,B., Wildgruber,D., and Vuilleumier,P. (2011). Emotional voice areas: anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. Cerebral Cortex bhr113.

Ethofer,T., Wiethoff,S., Anders,S., Kreifelts,B., Grodd,W., and Wildgruber,D. (2007). The voices of seduction: cross-gender effects in processing of erotic prosody. Social cognitive and affective neuroscience *2*, 334-337.

Faber,A., Gade,M., Negri,G., Lunardelli,A., Baumgaertner,A., Binkofski,F., Danek,A., and Rumiati,R.I. " Tomato and tuna" for testing action understanding after middle cerebral artery infarcts. JOURNAL OF NEUROLOGY 255, 47. 2008. Dr Dietrich Steinkipff Verlag, Darmstadty, Germany.

Faglia,L., and Vignolo,L.A. (1990). A case of "crossed aphasia" in which the integrity of the left hemisphere is assessed by MRI. The Italian Journal of Neurological Sciences *11*, 51-55.

Fastenau,P., Denburg,N., and Hufford,B. (1999). Adult norms for the Rey-Osterrieth Complex Figure Test and for supplemental recognition and matching trials from the Extended Complex Figure Test. The Clinical Neuropsychologist *13*, 30-47.

Fay,M. (2010). Confidence intervals that match Fisher's exact or Blaker's exact tests. Biostatistics *11*, 373-374.

Fecteau,S., Armony,J., Joanette,Y., and Belin,P. (2004). Is voice processing species-specific in human auditory cortex? An fMRI study. NeuroImage *23*, 840-848.

Fecteau,S., Armony,J.L., Joanette,Y., and Belin,P. (2005). Sensitivity to voice in human prefrontal cortex. Journal of neurophysiology *94*, 2251-2254.

Fiebach,C., Friederici,A., Muller,K., and Cramon,D. (2002). fMRI evidence for dual routes to the mental lexicon in visual word recognition. Journal of Cognitive Neuroscience *14*, 11-23.

Fiez,J., Raichle,M., Miezin,F., Petersen,S., Tallal,P., and Katz,W. (1995). PET studies of auditory and phonological processing: Effects of stimulus characteristics and task demands. Journal of Cognitive Neuroscience *7*, 357-375.

Fink,R., Brecher,A., Schwartz,M., and Robey,R. (2002). A computer-implemented protocol for treatment of naming disorders: Evaluation of clinician-guided and partially self-guided instruction. Aphasiology *16*, 1061-1086.

Fitch,W. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. The Journal of the Acoustical Society of America *102*, 1213-1222.

Fleming,D., Giordano,B.L., Caldara,R., and Belin,P. (2014). A language-familiarity effect for speaker discrimination without comprehension. Proceedings of the National Academy of Sciences *111*, 13795-13798.

Foulkes,P., and Docherty,G. (2006). The social life of phonetics and phonology. Journal of Phonetics *34*, 409-438.

Fowler,C. (1986). An event approach to the study of speech perception from a direct-realist perspective. Journal of Phonetics *14*, 3-28.

Fowler,C., and Dekle,D. (1991). Listening with eye and hand: Cross-modal contributions to speech perception. Journal of Experimental Psychology: Human Perception and Performance *17*, 816-828.

Fowler,C., and Rosenblum,L. (1990). Duplex perception: A comparison of monosyllables and slamming doors. Journal of Experimental Psychology: Human Perception and Performance *16*, 742-754.

Fridriksson,J., Bonilha,L., and Rorden,C. (2007). Severe Broca's aphasia without Broca's area damage. Behavioural neurology *18*, 237-238.

Fridriksson,J., Guo,D., Fillmore,P., Holland,A., and Rorden,C. (2013). Damage to the anterior arcuate fasciculus predicts non-fluent speech production in aphasia. Brain *136*, 3451-3460.

Friston,K., Holmes,A., Worsley,K., Poline,J., Frith,C., and Frackowiak,R. (1994). Statistical parametric maps in functional imaging: a general linear approach. Human Brain Mapping *2*, 189-210.

Galantucci,B., Fowler,C., and Turvey,M. (2006). The motor theory of speech perception reviewed. Psychonomic bulletin & review *13*, 361-377.

Gallese,V., Fadiga,L., Fogassi,L., and Rizzolatti,G. (1996). Action recognition in the premotor cortex. Brain *119*, 593-609.

Ganong,W. (1980). Phonetic categorization in auditory word perception. Journal of Experimental Psychology: Human Perception and Performance *6*, 110-125.

Garrett,M. (1975). The analysis of sentence production. Psychology of learning and motivation *9*, 133-177.

Garrido,L., Eisner,F., McGettigan,C., Stewart,L., Sauter,D., Hanley,J., Schweinberger,S., Warren,J., and Duchaine,B. (2009). Developmental phonagnosia: a selective deficit of vocal identity recognition. Neuropsychologia *47*, 123-131.

Gauthier,I., and Bukach,C. (2007). Should we reject the expertise hypothesis? Cognition *103,* 322-330.

Gauthier,L., Dehaut,F., and Joannette,Y. (1989). The Bells Test: a quantitative and qualitative test for visual neglect. International journal of clinical neuropsychology *4*, 155-159.

Gelfer,M., and Mikos,V. (2005). The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels. Journal of Voice *19*, 544-554.

George,M., Parekh,P., Rosinsky,N., Ketter,T., Kimbrell,T., Heilman,K., Herscovitch,P., and Post,R. (1996). Understanding emotional prosody activates right hemisphere regions. Arch Neurol *53*, 665-670.

Gerrits,E., and Schouten,M. (2004). Categorical perception depends on the discrimination task. Perception & Psychophysics *66*, 363-376.

Geshwind,N. (1965). Disconnection Syndromes in Man: Part I and II. Brain *88*, 237-294.

Geusebroek,J., Burghouts,G., and Smeulders,A. (2005). The Amsterdam library of object images. International Journal of Computer Vision *61*, 103-112.

Ghazanfar,A., and Rendall,D. (2008). Evolution of human vocal production. Current Biology *18*, R457-R460.

Gibson,Q. (1966). Is There a Problem About Appearances? The Philosophical Quarterly 319-328.

Gillebert,C., Mantini,D., Thijs,V., Sunaert,S., Dupont,P., and Vandenberghe,R. (2011). Lesion evidence for the critical role of the intraparietal sulcus in spatial attention. Brain awr085.

Giovagnoli,A. (1993). Crossed aphasia: Report of a rare case in a glioblastoma patient. The Italian Journal of Neurological Sciences *14*, 329-332.

Godecke,E., Hird,K., Lalor,E., Rai,T., and Phillips,M. (2012). Very early poststroke aphasia therapy: A pilot randomized controlled efficacy trial. international Journal of Stroke *7*, 635-644.

Godefroy,O. (2013). The Behavioral and Cognitive Neurology of Stroke (Cambridge: Cambridge University Press).

Goldinger,S. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. Journal of Experimental Psychology: Learning, Memory and Cognition *22*, 1166.

Goldinger,S. A complementary-systems approach to abstract and episodic speech perception. Proceedings of the 16th international congress of phonetic sciences , 49-54. 2007.

Goodglass,H., Kaplan,E., and Barresi,B. (2001). The assessment of aphasia and related disorders (USA: Lippincott Williams & Wilkins).

Goodglass,H., and Quadfasel,F. (1954). Language laterality in left-handed aphasics. Brain *77*, 521-548.

Goodman-Schulman,R., and Caramazza,A. (1987). Patterns of dysgraphia and the nonlexical spelling process. Cortex *23*, 143-148.

Gordon,J. (1998). The fluency dimension in aphasia. Aphasiology *12*, 673-688.

Gorgolewski,K., Storkey,A., Bastin,M., and Pernet,C. (2012). Adaptive thresholding for reliable topological inference in single subject fMRI analysis. Frontiers in Human Neuroscience *6*.

Gorgolewski,K., Storkey,A., Bastin,M., Whittle,I., and Pernet,C. (2013a). Single subject fMRI test-retest reliability metrics and confounding factors. NeuroImage *69*, 231-243.

Gorgolewski,K., Storkey,A., Bastin,M., Whittle,I., Wardlaw,J., and Pernet,C. (2013b). A test-retest fMRI dataset for motor, language and spatial attention functions. GigaScience *2*, 1-4.

Graham,K., Lambon Ralph,M., and Hodges,J. (1999a). A questionable semantics: The interaction between semantic knowledge and autobiographical experience in semantic dementia. Cognitive neuropsychology *16*, 689-698.

Graham,K., Patterson,K., Pratt,K., and Hodges,J. (1999b). Relearning and subsequent forgetting of semantic category exemplars in a case of semantic dementia. Neuropsychology *13*, 359.

Grosjean,F. (1980). Spoken word recognition processes and the gating paradigm. Perception & Psychophysics *28*, 267-283.

Gross,R., and Grossman,M. (2008). Update on apraxia. Current neurology and neuroscience reports *8*, 490-496.

Grossberg,S. (1980). Direct perception or adaptive resonance? Behavioral and Brain Sciences *3*, 385-386.

Gubbay,S., and De Klerk,N. (1995). A study and review of developmental dysgraphia in relation to acquired dysgraphia. Brain and Development *17*, 1-8.

Guerreiro,M., Castrocaldas,A., and Martins,I.P. (1995). Aphasia following right hemisphere lesion in a woman with left hemisphere injury in childhood. Brain and Language *49*, 280-288.

Ha,J., Pyun,S., Hwang,Y., and Sim,H. (2012). Lateralization of cognitive functions in aphasia after right brain damage. Yonsei medical journal *53*, 486-494.

Haaland,K., Harrington,D., and Knight,R. (2000). Neural representations of skilled movement. Brain *123*, 2306-2313.

Haaland,K., and Miranda,F. (1982). Psychometric and CT scan measurements in a case of crossed aphasia in a dextral. Brain and Language *17*, 240-260.

Habib,M. (2000). The neurological basis of developmental dyslexia an overview and working hypothesis. Brain *123*, 2373-2399.

Habib,M., Joanette,Y., Ali Cherif,A., and Poncet,M. (1983). Crossed aphasia in dextrals: a case report with special reference to site of lesion. Neuropsychologia *21*, 413-418.

Hackett,T., Stepniewska,I., and Kaas,J. (1998). Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. Journal of Comparative Neurology *394*, 475-495.

Hagoort,P. (2013). MUC (Memory, Unification, Control) and beyond. Frontiers in Psychology *4*.

Hailstone,J., Crutch,S., Vestergaard,M., Patterson,R., and Warren,J. (2010a). Progressive associative phonagnosia: a neuropsychological analysis. Neuropsychologia *48*, 1104-1114.

Hailstone,J., Crutch,S., and Warren,J. (2010b). Voice recognition in dementia. Behavioural neurology *23*, 163-164.

Hall,J., and Guyton,A. (2005). Textbook of Medical Physiology (Philadelphia: Elsevier).

Hanson,H., and Chuang,E. (1999). Glottal characteristics of male speakers: acoustic correlates and comparison with female data. The Journal of the Acoustical Society of America *106*, 1064-1077.

Hatfield,F., and Howard,D. (1978). Linguistic Investigations of Aphasia (London: College of Speech and Language Therapists).

Hauk,O., Johnsrude,I., and Pulvermuller,F. (2004). Somatotopic representation of action words in human motor and premotor cortex. Neuron *41*, 301-307.

Haxby,J., Gobbini,M., Furey,M., Ishai,A., Schouten,J., and Pietrini,P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science *293*, 2425-2430.

Heilman,K., Bowers,D., Speedie,L., and Coslett,H. (1984). Comprehension of affective and nonaffective prosody. Neurology *34*, 917.

Heiss,W., and Thiel,A. (2006). A proposed regional hierarchy in recovery of post-stroke aphasia. Brain and Language *98*, 118-123.

Helm-Estabrooks,N. (2002). Cognition and aphasia: A discussion and a study. Journal of Communication Disorders *35*, 171-186.

Henderson,V. (1983). Speech fluency in crossed aphasia. Brain *106*, 837-857.

Herman,A., Houde,J., Vinogradov,S., and Nagarajan,S. (2013). Parsing the phonological loop: activation timing in the dorsal speech stream determines accuracy in speech reproduction. The Journal of Neuroscience *33*, 5439-5453.

Herrmann,M., and Wallesch,C. (1990). Expectations of psychosocial adjustment in aphasia: a MAUT study with the Code-Muller Scale of Psychosocial Adjustment. Aphasiology *4*, 527-538.

Hewlett,N., and Beck,J. (2006). An introduction to the science of phonetics (New Jersey: Lawrence Erlbaum Associates Inc.).

Hickok,G., and Poeppel,D. (2000). Towards a functional neuroanatomy of speech perception. Trends in Cognitive Sciences *4*, 131-138.

Hickok,G. (2009). The functional neuroanatomy of language. Physics of life reviews *6*, 121-143.

Hickok,G. (2010). The role of mirror neurons in speech perception and action word semantics. Language and cognitive processes *25*, 749-776.

Hickok,G., and Poeppel,D. (2007). The cortical organization of speech processing. Nature Reviews Neuroscience *8*, 393-402.

Hillenbrand,J., Clark,M., and Nearey,T. (2001). Effects of consonant environment on vowel formant patterns. The Journal of the Acoustical Society of America *109*, 748-763.

Hillis,A., Tuffiash,E., Wityk,R., and Barker,P. (2002). Regions of neural dysfunction associated with impaired naming of actions and objects in acute stroke. Cognitive Neuropsychology, *19*, 523-534.

Hillis,A. (2007). Aphasia progress in the last quarter of a century. Neurology *69*, 200-213.

Hillis,A., Work,M., Barker,P., Jacobs,M., Breese,E., and Maurer,K. (2004). Re-examining the brain regions crucial for orchestrating speech articulation. Brain *127*, 1479-1487.

Hinckley,J. (1998). Investigating the predictors of lifestyle satisfaction among younger adults with chronic aphasia. Aphasiology *12*, 509-518.

Hinckley,J. (2002). Vocational and social outcomes of adults with chronic aphasia. Journal of Communication Disorders *35*, 543-560.

Hintze,J., and Nelson,R. (1998). Violin plots: a box plot-density trace synergism. The American Statistician *52*, 181-184.

Hirono,N., Mori,E., Ishii,K., Imamura,T., Tanimukai,S., Kazui,H., Hashimoto,M., Takatsuki,Y., Kitagaki,H., and Sasaki,M. (2000). Neuronal substrates for semantic memory: a positron emission tomography study in Alzheimer's disease. Dementia And Geriatric Cognitive Disorders *12*, 15-21.

Hoffstaedter,F., Grefkes,C., Caspers,S., Roski,C., Fox,P., Zilles,K., and Eickhoff,S. (2012). Functional connectivity of the mid-cingulate cortex. Klinische Neurophysiologie *43*, 128.

Holland,A., Frattali,C., and Fromm,D. (1999). Communication activities of daily living: CADL-2 (Baltimore: University Park Press).

Holt,L., and Lotto,A. (2010). Speech perception as categorization. Attention, Perception & Psychophysics *72*, 1218-1227.

Holt,L., Lotto,A., and Kluender,K. (1998). Spectral contrast in perception of VCV syllables. The Journal of the Acoustical Society of America *104*, 1759.

Hopfield,J. (1982). Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences *79*, 2554-2558.

Hornak,J., Rolls,E., and Wade,D. (1996). Face and voice expression identification in patients with emotional and behavioural changes following ventral frontal lobe damage. Neuropsychologia *34*, 247-261.

Howard,D., and Patterson,K. (1992). The Pyramids and Palm Trees Test: A test of semantic access from words and pictures (Bury St. Edmunds: Thames Valley Test Company).

Intercollegiate Stroke Working Party. National clinical guideline for stroke, 4th edition. 97. 2012. London, Royal College of Physicians.

Ishizaki,M., Ueyama,H., Nishida,Y., Imamura,S., Hirano,T., and Uchino,M. (2012). Crossed aphasia following an infarction in the right corpus callosum. Clinical neurology and neurosurgery *114*, 161-165.

Ito,T., Tiede,M., and Ostry,D. (2009). Somatosensory function in speech perception. Proceedings of the National Academy of Sciences *106*, 1245-1248.

Jackson,J. (1864). Hemiplegia on the right side, with loss of speech. British Medical Journal *1*, 572.

Jancke,L., Wustenberg,T., Scheich,H., and Heinze,H.J. (2002). Phonetic perception and the temporal cortex. NeuroImage *15*, 733-746.

Jenkinson,M., Beckmann,C., Behrens,T., Woolrich,M., and Smith,S. (2012). FSL. NeuroImage *62*, 782-790.

Johnson,K. (1990). The role of perceived speaker identity in F0 normalization of vowels. Journal of the Acoustical Society of America *88*, 642-654.

Johnson,K. (1997). Speech perception without speaker normalization: An exemplar model. In Talker Variability in Speech Processing, K. Johnson, and J. Mullennix, eds. (San Diego: Academic Press), pp. 145-165.

Johnson,K., Ladefoged,P., and Lindau,M. (1993). Individual differences in vowel production. The Journal of the Acoustical Society of America *94*, 701-714.

Johnson,K., and Mullennix,J. (1997). Talker Variability in Speech Processing (London: Academic Press).

Johnstone,T., Van Reekum,C., Oakes,T., and Davidson,R. (2006). The voice of emotion: an FMRI study of neural responses to angry and happy vocal expressions. Social cognitive and affective neuroscience *1*, 242-249.

Jones,A., Farrall,A., Belin,P., and Pernet,C. (2015). Hemispheric association and dissociation of voice and speech information processing in stroke. Cortex *71*, 232-239.

Jones,E., Dell'Anna,M., Molinari,M., Rausell,E., and Hashikawa,T. (1995). Subdivisions of macaque monkey auditory cortex revealed by calcium-binding protein immunoreactivity. Journal of Comparative Neurology *362*, 153-170.

Kaas,J., and Hackett,T. (1999). 'What' and 'where' processing in auditory cortex. Nat Neurosci *2*, 1045-1047.

Kaas,J., and Hackett,T. (2000). Subdivisions of auditory cortex and processing streams in primates. Proceedings of the National Academy of Sciences *97*, 11793-11799.

Kaganovich,N., Francis,A., and Melara,R. (2006). Electrophysiological evidence for early interaction between talker and linguistic information during speech perception. Brain research *1114*, 161-172.

Karnath,H., Berger,M., Kuker,W., and Rorden,C. (2004). The anatomy of spatial neglect based on voxelwise statistical analysis: a study of 140 patients. Cerebral Cortex *14*, 1164-1172.

Karnath,H., and Rorden,C. (2012). The anatomy of spatial neglect. Neuropsychologia *50*, 1010-1017.

Kashino,M. (2006). Phonemic restoration: The brain creates missing speech sounds. Acoustical science and technology *27*, 318-321.

Kasselimis,D., Simos,P., Economou,A., Peppas,C., Evdokimidis,I., and Potagas,C. (2013). Are memory deficits dependent on the presence of aphasia in left brain damaged patients? Neuropsychologia *51*, 1773-1776.

Kawahara,H. Exemplar-based voice quality analysis and control using a high quality auditory morphing procedure based on straight. ISCA Tutorial and Research Workshop on Voice Quality: Functions, Analysis and Synthesis . 2003.

Kawahara,H. (2006). STRAIGHT, exploitation of the other aspect of VOCODER: Perceptually isomorphic decomposition of speech sounds. Acoustical science and technology *27*, 349-353.

Kawahara,H., and Akahane-Yamada,R. (2006). STRAIGHT as a research tool for L2 study: How to manipulate segmental and supra-segmental features. The Journal of the Acoustical Society of America *120*, 3137.

Kay,J., Lesser,R., and Coltheart,M. (1992). PALPA: Psycholinguistic Assessment of Language Performance in Aphasia (London: Lawrence Erlbaum).

Kearns,K. (1993). Functional Outcome: Methodological considerations. Clinical aphasiology *21*, 67-72.

Kelly,H., Brady,C., and Enderby,P. Speech and Language therapy for aphasia following stroke. Cochrane Database Systematic Review 5. 2010.

Kertesz,A. (1979). Aphasia and associated disorders: Taxonomy, localization, and recovery (New York: Grune & Stratton).

Kertesz,A. (1982). Western Aphasia Battery (New York: Grune & Stratton).

Kim,W., Yang,E., and Paik,N. (2013). Neural Substrate Responsible for Crossed Aphasia. Journal of Korean medical science *28*, 1529-1533.

Kingston,J., and Diehl,R. (1995). Intermediate properties in the perception of distinctive feature values. Papers in laboratory phonology *4*, 7-27.

Klatt,D. (1980). Software for a cascade/parallel formant synthesizer. The Journal of the Acoustical Society of America *67*, 971-995.

Kleiner,M., Brainard,D., Pelli,D., Ingling,A., Murray,R., and Broussard,C. (2007). What's new in Psychtoolbox-3. Perception *36*, 1.

Kluender,K., Diehl,R., and Killeen,P. (1987). Japanese quail can learn phonetic categories. Science *237*, 1195-1197.

Kluender,K., and Lotto,A. (1994). Effects of first formant onset frequency on [-voice] judgments result from auditory processes not specific to humans. The Journal of the Acoustical Society of America *95*, 1044-1052.

Kluender,K., Lotto,A., and Holt,L. (2006). Contributions of nonhuman animal models to understanding human speech perception. Listening to speech: An auditory perspective 203-220.

Knosche,T., Lattner,S., Maess,B., Schauer,M., and Friederici,A. (2002). Early parallel processing of auditory word and voice information. NeuroImage *17*, 1493-1503.

Kobayashi,S., and Ugawa,Y. (2013). Relationships between Aphasia and Apraxia. Journal of Neurology & Translational Neuroscience *2*, 1028.

Koch,D., McGee,T., Bradlow,A., and Kraus,N. (1999). Acoustic-phonetic approach toward understanding neural processes and speech perception. American Academy of Neurology *10*, 304-318.

Kotz,S., Cappa,S., von Cramon,D., and Friederici,A. (2002). Modulation of the lexical-semantic network by auditory semantic priming: An event-related functional MRI study. NeuroImage *17*, 1761-1772.

Kreisler,A., Godefroy,O., Delmaire,C., Debachy,B., Leclercq,M., Pruvo,J., and Leys,D. (2000). The anatomy of aphasia revisited. Neurology *54*, 1117-1123.

Kreitewolf,J., Gaudrain,E., and von Kriegstein,K. (2014). A neural mechanism for recognizing speech spoken by different speakers. NeuroImage *91*, 375-385.

Kruschke,J. (1992). ALCOVE: an exemplar-based connectionist model of category learning. Psychological Review *99*, 22.

Kuhl,P., and Miller,J. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. The Journal of the Acoustical Society of America *63*, 905-917.

Kuhl,P., and Padden,D. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. The Journal of the Acoustical Society of America *73*, 1003-1010.

Kurthen,M., Linke,D., Elger,C., and Schramm,J. (1992). Linguistic perseveration in dominant-side intracarotid amobarbital tests. Cortex *28*, 209-219.

Laganaro,M., and Alario,F. (2006). On the locus of the syllable frequency effect in speech production. Journal of Memory and Language *55*, 178-196.

Lahiri,A., and Reetz,H. (2002). Underspecified recognition. Laboratory phonology *7*, 637-676.

Laing,E., Liu,R., Lotto,A., and Holt,L. (2012). Tuned with a tune: talker normalization via general auditory processes. Frontiers in Psychology *3*, 203.

Landis,T., Buttet,J., Assal,G., and Graves,R. (1982). Dissociation of ear preference in monaural word and voice recognition. Neuropsychologia *20*, 501-504.

LaPointe,L. (2005). Aphasia and related neurogenic language disorders (UK: Thieme Medical).

Lass,N., Hughes,K., Bowyer,M., Waters,L., and Bourne,V. (1976). Speaker sex identification from voiced, whispered, and filtered isolated vowels. The Journal of the Acoustical Society of America *59*, 675-678.

Latimer,N., Dixon,S., and Palmer,R. (2013). Cost-utility of self-managed computer therapy for people with aphasia. International journal of technology assessment in health care *29*, 402-409.

Latinus,M., Crabbe,F., and Belin,P. (2009). fMRI investigations of voice identity perception. NeuroImage *47*, S156.

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

Law,J., Pringle,A., Irving,A., Huby,G., Smith,M., Conochie,D., Haworth,C., and Burston,A. Aphasia in Scotland Project. 3. 2007. Centre for Integrated Healthcare Research.

Lazar,R., Marshall,R., Pile-Spellman,J., Duong,H., Mohr,J., Young,W., Solomon,R., Perera,G., and DeLaPaz,R. (2000). Interhemispheric transfer of language in patients with left frontal cerebral arteriovenous malformation. Neuropsychologia *38*, 1325-1332.

Lebel,C., and Beaulieu,C. (2009). Lateralization of the arcuate fasciculus from childhood to adulthood and its relation to cognitive abilities in children. Human Brain Mapping *30*, 3563-3573.

Leff,A., Schofield,T., Crinion,J., Seghier,M., Grogan,A., Green,D., and Price,C. (2009). The left superior temporal gyrus is a shared substrate for auditory short-term memory and speech comprehension: evidence from 210 patients with stroke. Brain *132*, 3401-3410.

Lessa Mansur,L., Radanovic,M., Santos Penha,S., Iracema Zanotto de Mendonoa,L., and Cristina Adda,C. (2006). Language and visuospatial impairment in a case of crossed aphasia. Laterality *11*, 525-539.

Levitt,H. (1971). Transformed up and down methods in psychoacoustics. The Journal of the Acoustical Society of America *49*, 467-477.

Lew,H., Jerger,J., Guillory,S., and Henry,J. (2007). Auditory dysfunction in traumatic brain injury. Journal of rehabilitation research and development *44*, 921.

Lezak,M., Howieson D., Loring D., Hannay H., and Fischer J. (2004). Neuropsychological Assessment (Oxford: Oxford University Press).

Liberman,A., Cooper,F., Shankweiler,D., and Studdert-Kennedy,M. (1967). Perception of the speech code. Psychological Review *74*, 431-461.

Liberman,A., Delattre,P., and Cooper,F. (1952). The role of selected stimulus-variables in the perception of the unvoiced stop consonants. The American journal of psychology 497-516.

Liberman,A., and Mattingly,I. (1985). The motor theory of speech perception revised. Cognition *21*, 1-36.

Liberman,A., and Mattingly,I. (1989). A specialization for speech perception. Science *243*, 489-494.

Liberman,A., and Whalen,D. (2000). On the relation of speech to language. Trends in Cognitive Sciences *4*, 187-196.

Lichtheim,L. (1885). On aphasia. In Broca's Region, Y. Grodzinsky, and K. Amunts, eds. (Oxford: Oxford University Press), pp. 318-347.

Liebenthal,E., Desai,R., Ellingson,M., Ramachandran,B., Desai,A., and Binder,J. (2010). Specialization along the left superior temporal sulcus for auditory categorization. Cerebral Cortex *20*, 2958-2970.

Lightfoot,D. (1989). The child's trigger experience: Degree-0 learnability. Behavioral and Brain Sciences *12*, 321-375.

Likert,R. (1932). A technique for the measurement of attitudes. Archives of psychology *22*, 5-55.

Lincoln N, McGuirk E, Mulley G, Lendrem,W., Jones A, and Mitchell,J. (1984). Effectiveness of speech therapy for aphasic stroke patients. A randomised controlled trial. The Lancet *323*, 1197-1200.

Lisker,L., and Abramson,A. (1963). Crosslanguage study of voicing in initial stops. The Journal of the Acoustical Society of America *35*, 1889-1890.

Lotto,A. (2000). Language acquisition as complex category formation. Phonetica *57*, 189-196.

Love,R., and Webb,W. (1996). Neurology for the Speech-Language Pathologist (Boston: Butterworth-heinemann).

Luria,A. (1966). Higher Cortical Functions in Man. (New York: Basic Books).

Macmillan,N., and Creelman,C. (2004). Detection theory: A user's guide (London: Erlbaum Associates).

Maesto,F., Saldada,C., Amo,C., Gonzaílez-Hidalgo,M., Fernandez,A., Fernandez,S., Mata,P., Papanicolaou,A., and Ortiz,T. (2004). Can small lesions induce language reorganization as large lesions do? Brain and Language *89*, 433-438.

Marie,P. (1906). Revision de la question de l'aphasie: L'aphasie de 1861 a 1866: Essai de critique historique sur la genese de la doctrine de Broca. Semaine medicale 565-571.

Marien,P., Engelborghs,S., Vignolo,L., and De Deyn,P. (2001). The many faces of crossed aphasia in dextrals: report of nine cases and review of the literature. European Journal of Neurology *8*, 643-658.

Marien,P., Paghera,B., De Deyn,P., and Vignolo,L. (2004). Adult crossed aphasia in dextrals revisited. Cortex *40*, 41-74.

Marshall,J., and Halligan,P. (1992). Crossed aphasia in a dextral without "minor" hemisphere signs. Behavioural neurology *5*, 247-250.

Marshall,R., Basilakos,A., and Love-Myers,K. (2013). Further evidence of auditory extinction in aphasia. J Speech Lang Hear Res *56*, 236-249.

Marslen-Wilson,W. (1987). Functional parallelism in spoken word-recognition. Cognition *25*, 71-102.

Martin,R., Lesch,M., and Bartha,M.C. (1999). Independence of input and output phonology in word processing and short-term memory. Journal of Memory and Language *41*, 3-29.

Massaro,D. (1989). Testing between the TRACE model and the fuzzy logical model of speech perception. Cognitive psychology *21*, 398-421.

Massaro,D., and Oden,G. (1980). Evaluation and integration of acoustic features in speech perception. The Journal of the Acoustical Society of America *67*, 996-1013.

Mastronardi,L., Ferrante,L., Maleci,A., Puzzilli,F., Lunardi,P., and Schettini,G. (1994). Crossed aphasia. An update. Neurosurgical review *17*, 299-304.

Mazoyer,B., Zago,L., Jobard,G., Crivello,F., Joliot,M., Perchey,G., Mellet,E., Petit,L., and Tzourio-Mazoyer,N. (2014). Gaussian mixture modeling of hemispheric lateralization for language in a large sample of healthy individuals balanced for handedness. PLoS ONE *9*, e101165.

Mazziotta,J., Toga,A., Evans,A., Fox,P., Lancaster,J., Zilles,K., Woods,R., Paus,T., Simpson,G., and Pike,B. (2001). A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). Philosophical Transactions of the Royal Society B: Biological Sciences *356*, 1293-1322.

Mazzocchi,F., and Vignolo,L. (1979). Localisation of lesions in aphasia: Clinical-CT scan correlations in stroke patients. Cortex *15*, 627-653.

McClelland,J., and Elman,J. (1986). The TRACE model of speech perception. Cognitive psychology *18*, 1-86.

McGettigan,C., and Scott,S. (2012). Cortical asymmetries in speech perception: what's wrong, what's right and what's left? Trends in Cognitive Sciences *16*, 269-276.

McGurk,H., and MacDonald,J. (1976). Hearing lips and seeing voices. Nature *264*, 746-748.

McKone,E., Kanwisher,N., and Duchaine,B. (2007). Can generic expertise explain special processing for faces? Trends in Cognitive Sciences *11*, 8-15.

McLachlan,N., and Wilson,S. (2010). The central role of recognition in auditory perception: a neurobiological model. Psychological Review *117*, 175.

McNeil,M., and Pratt,S. (2001). Defining aphasia: Some theoretical and clinical implications of operating from a formal definition. Aphasiology *15*, 901-911.

McQueen,J., Cutler,A., and Norris,D. (2006). Phonological abstraction in the mental lexicon. Cognitive Science *30*, 1113-1126.

Mei,L., Xue,G., Lu,Z., He,Q., Zhang,M., Wei,M., Xue,F., Chen,C., and Dong,Q. (2014). Artificial language training reveals the neural substrates underlying addressed and assembled phonologies. PLoS ONE *9*.

Meinzer,M., Djundja,D., Barthel,G., Elbert,T., and Rockstroh,B. (2005). Long-Term Stability of Improved Language Functions in Chronic Aphasia After Constraint-Induced Aphasia Therapy. Stroke *36*, 1462-1466.

Meister,I., Wilson,S., Deblieck,C., Wu,A., and Lacoboni,M. (2007). The essential role of premotor cortex in speech perception. Curr. Biol. *17*, 1692-1696.

Mesulam,M., Thompson,C., Weintraub,S., and Rogalski,E. (2015). The Wernicke conundrum and the anatomy of language comprehension in primary progressive aphasia. Brain *138*, 2423-2437.

Micceri,T. (1989). The unicorn, the normal curve, and other improbable creatures. Psychological Bulletin *105*, 156.

Miceli,G., Silveri,M., and Caramazza,A. (1985). Cognitive analysis of a case of pure dysgraphia. Brain and Language *25*, 187-212.

Miller,C., and Wagstaff,D. (2011). Behavioral profiles associated with auditory processing disorder and specific language impairment. Journal of Communication Disorders *44*, 745-763.

Miller,G., and Johnson-Laird,P. (1976). Language and perception (Cambridge: Belknap Press), pp. viii, 760.

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

Mitterer,H., Yoneyama,K., and Ernestus,M. (2008). How we hear what is hardly there: Mechanisms underlying compensation for/t/-reduction in speech comprehension. Journal of Memory and Language *59*, 133-152.

Moeller,K., Willmes,K., and Klein,E. (2015). A review on functional and structural brain connectivity in numerical cognition. Frontiers in Human Neuroscience *9*, 227.

Molenberghs,P., Cunnington,R., and Mattingley,J. (2012). Brain regions with mirror properties: a meta-analysis of 125 human fMRI studies. Neuroscience & Biobehavioral Reviews *36*, 341-349.

Morosan,P., Rademacher,J., Schleicher,A., Amunts,K., Schormann,T., and Zilles,K. (2001). Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. NeuroImage *13*, 684-701.

Morris,J., and Franklin,S. (1995). Assessment and remediation of a speech discrimination deficit in a dysphasic patient. In Case studies in clinical linguistics, M. Perkins, and S. Howard, eds. (London: Whurr Publishers), pp. 245-270.

Morris,J., Scott,S., and Dolan,R. (1999). Saying it with feeling: neural responses to emotional vocalizations. Neuropsychologia *37*, 1155-1163.

Mort,D., Malhotra,P., Mannan,S., Rorden,C., Pambakian,A., Kennard,C., and Husain,M. (2003). The anatomy of visual neglect. Brain *126*, 1986-1997.

Mortley,J., Wade,J., and Enderby,P. (2004). Superhighway to promoting a client-therapist partnership? Using the Internet to deliver word-retrieval computer therapy, monitored remotely with minimal speech and language therapy input. Aphasiology *18*, 193-211.

Morton,J., and Patterson,K. (1980). A new attempt at an interpretation, or, an attempt at a new interpretation. Deep dyslexia 91-118.

Mottonen,R., Calvert,G., Jaaskelainen,L., Matthews,P., Thesen,T., Tuomainen,J., and Sams,M. (2006). Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. NeuroImage *30*, 563-569.

Mullennix,J., Johnson,K., Topcu-Durgun,M., and Farnsworth,L. (1995). The perceptual representation of voice gender. The Journal of the Acoustical Society of America *98*, 3080-3095.

Mullennix,J., and Pisoni,D. (1990). Stimulus variability and processing dependencies in speech perception. Perception & Psychophysics *47*, 379-390.

Muller,N., and Knight,R. (2006). The functional neuroanatomy of working memory: contributions of human brain lesion studies. Neuroscience *139*, 51-58.

Mummery,C., Ashburner,J., Scott,S., and Wise,R. (1999). Functional neuroimaging of speech perception in six normal and two aphasic subjects. The Journal of the Acoustical Society of America *106*, 449-457.

Murray,L. (2012). Attention and other cognitive deficits in aphasia: Presence and relation to language and communication measures. Am J Speech Lang Pathol *21*, S51-S64.

Musiek,F., Mohanani,A., Wierzbinski,E., Kilgore,G., Hunter,J., and Marotto,J. (2011). The non-classical pathway: Too great to be ignored. The Hearing Journal *64*, 6-8.

Newhart,M., Ken,L., Kleinman,J., Heidler-Gary,J., and Hillis,A. (2007). Neural networks essential for naming and word comprehension. Cognitive and Behavioral Neurology *20*, 25-30.

Nguyen,N., Wauquier,S., and Tuller,B. (2009). The Dynamical Approach to Speech Perception: From fine phonetic detail to abstract phonological categories. In Approaches to Phonological Complexity, F. Pellegrinio, E. Marsico, I. Chitoran, and C. Coupe, eds. (Germany: Mouton de Gruyter), pp. 193-218.

Nickels,L., Howard,D., and Best,W. (1997). Fractionating the articulatory loop: Dissociations and associations in phonological recoding in aphasia. Brain and Language *56*, 161-182.

Norris,D. (1994). Shortlist: A connectionist model of continuous speech recognition. Cognition *52*, 189-234.

Norris,D. (1999). The merge model: Speech perception is bottom-up. The Journal of the Acoustical Society of America *106*, 2295.

Norris,D., and McQueen,J. (2008). Shortlist B: A Bayesian model of continuous speech recognition. Psychological Review *115*, 357.

Norris,D., McQueen,J., and Cutler,A. (2000). Merging information in speech recognition: Feedback is never necessary. Behavioral and Brain Sciences *23*, 299-325.

Norris,D., McQueen,J., Cutler,A., Butterfield,S., and Kearns,R. (2001). Language-universal constraints on speech segmentation. Language and cognitive processes *16*, 637-660.

Nygaard,L., and Pisoni,D. (1996). Learning voices. The Journal of the Acoustical Society of America *99*, 2589-2603.

Nygaard,L., Sommers,M., and Pisoni,D. (1994). Speech Perception as a Talker-Contingent Process. Psychological Science *5*, 42-46.

Ocklenburg,S., Beste,C., Arning,L., Peterburs,J., and Gunturkun,O. (2014). The ontogenesis of language lateralization and its relation to handedness. Neuroscience & Biobehavioral Reviews *43*, 191-198.

Oldfield,R. (1971). The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia *9*, 97-113.

Omar,R., Hailstone,J., Warren,J., Crutch,S., and Warren,J. (2010). The cognitive organization of music knowledge: a clinical analysis. Brain *133*, 1200-1213.

Oppenheim,H. (1891). Zur Pathologie der Grosshirngeschwulste. European Archives of Psychiatry and Clinical Neuroscience *22*, 27-72.

Osmon,D., Panos,J., Kautz,P., and Gandhavadi,B. (1998). Crossed aphasia in a dextral: a test of the Alexander-Annett theory of anomalous organization of brain function. Brain and Language *63*, 426-438.

Paghera,B., Marien,P., and Vignolo,L. (2003). Crossed aphasia with left spatial neglect and visual imperception: a case report. Neurological Sciences *23*, 317-322.

Pallier,C., Colome,A., and Sebastian-Galles,N. (2001). The Influence of Native-Language Phonology on Lexical Access: Exemplar-Based Versus Abstract Lexical Entries. Psychological Science *12*, 445-449.

Palmer,R., Enderby,P., Cooper,C., Latimer,N., Juliousm,S., Paterson,G., Dimairo,M., Dixon,S., Mortley,J., Hilton,R., Delaney,A., and Hughes,H. (2012). Computer Therapy Compared With Usual Care for People With Long-Standing Aphasia Poststroke A Pilot Randomized Controlled Trial. Stroke *43*, 1904-1911.

Palmeri,T., Goldinger,S., and Pisoni,D. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. Journal of experimental psychology: Learning, memory, and cognition *19*, 309.

Papagno,C., la Sala,S., and Basso,A. (1993). Ideomotor apraxia without aphasia and aphasia without apraxia: the anatomical support for a double dissociation. Journal of Neurology, Neurosurgery & Psychiatry *56*, 286-289.

Paparounas,K., Eftaxias,D., and Akritidis,N. (2002). Dissociated crossed aphasia: a challenging language representation disorder. Neurology *59*, 441-442.

Papathanasiou,I., Coppens,P., and Potagas,C. (2011). Aphasia and related neurogenic communication disorders (Malloy: Jones & Bartlett Publishers).

Parr,S. (2007). Living with severe aphasia: Tracking social exclusion. Aphasiology *21*, 98-123.

Parr,S., Byng,S., Gilpin,S., and Ireland,C. (1997). Talking about aphasia: Living with loss of language after stroke (Philadelphia: Open University Press).

Patidar,Y., Gupta,M., Khwaja,G., Chowdhury,D., Batra,A., and Dasgupta,A. (2013). A case of crossed aphasia with apraxia of speech. Annals of Indian Academy of Neurology *16*, 428.

Pazzaglia,M., Pizzamiglio,L., Pes,E., and Aglioti,S. (2008). The Sound of Actions in Apraxia. Current Biology *18*, 1766-1772.

Pelli,D. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial vision *10*, 437-442.

Peperkamp,S. (2003). Phonological Acquisition: Recent Attainments and New Challenges. Language and Speech *46*, 87-113.

Peretz,I., Kolinsky,R., Tramo,M., Labrecque,R., Hublet,C., Demeurisse,G., and Belleville,S. (1994). Functional dissociations following bilateral lesions of auditory cortex. Brain *117 (Pt 6)*, 1283-1301.

Pernet,C., McAleer,P., Latinus,M., Gorgolewski,K.J., Charest,I., Bestelmeyer,P.E., Watson,R., Fleming,D., Crabbe,F., Valdes-Sosa,M., and Belin,P. (2015a). The human Voice Areas: spatial organisation and inter-individual variability in temporal and extra-temporal cortices. Frontiers in Psychology *1*, 164-174.

Pernet,C. (2014). Misconceptions in the use of the General Linear Model applied to functional MRI: a tutorial for junior neuro-imagers. Frontiers in Neuroscience *8*.

Pernet,C., and Belin,P. (2012). Role of pitch and timber in gender perception. Frontiers in Psychology *3*, 1-11.

Pernet,C., Belin,P., and Jones,A. (2013). Behavioral evidence of a dissociation between voice gender categorization and phoneme categorization using auditory morphed stimuli. Frontiers in Psychology *4*, 1018.

Pernet,C., Charest,I., Belizaire,G., Zatorre,R., and Belin,P. (2007). The temporal voice areas: spatial characterization and variability. NeuroImage *36*, S1-S168.

Pernet,C., McAleer,P., Latinus,M., Gorgolewski,K., Charest,I., Bestelmeyer,P., Watson,R., Fleming,D., Crabbe,F., and Valdes-Sosa,M. (2015b). The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. NeuroImage *119*, 164-174.

Pernet,C., Wilcox,R., and Rousselet,G. (2012). Robust correlation analyses: false positive and power validation using a new open source Matlab toolbox. Frontiers in Psychology *3*.

Perrachione,T., and Wong,P. (2007). Learning to recognize speakers of a non-native language: Implications for the functional organization of human auditory cortex. Neuropsychologia *45*, 1899-1910.

Perreault,C., and Mathew,S. (2012). Dating the origin of language using phonemic diversity. PLoS ONE *7*, e35289.

Pisoni,D. (1977). Identification and discrimination of the relative onset time of two component tones: implications for voicing perception in stops. The Journal of the Acoustical Society of America *61*, 1352-1361.

Planton,S., Jucla,M., Roux,F., and Demonet,J.-F. (2013). The 'handwriting brain': a meta-analysis of neuroimaging studies of motor versus orthographic processes. Cortex *49*, 2772-2787.

Poeppel,D., and Hickok,G. (2004). Towards a new functional anatomy of language. Cognition *92*, 1-12.

Poldrack,R., Wagner,A., Prull,M., Desmond,J., Glover,G., and Gabrieli,J. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. NeuroImage *10*, 15-35.

Poremba,A., Saunders,R., Crane,A., Cook,M., Sokoloff,L., and Mishkin,M. (2003). Functional mapping of the primate auditory system. Science *299*, 568-572.

Posteraro,L., Zinelli,P., and Mazzucchi,A. (1988). Selective impairment of the graphemic buffer in acquired dysgraphia: A case study. Brain and Language *35*, 274-286.

Price,C., and Devlin,J. (2011). The interactive account of ventral occipitotemporal contributions to reading. Trends in Cognitive Sciences *15*, 246-253.

Rama,P., Martinkauppi,S., Linnankoski,I., Koivisto,J., Aronen,H., and Carlson,S. (2001). Working memory of identification of emotional vocal expressions: an fMRI study. NeuroImage *13*, 1090-1101.

Rama,P., Poremba,A., Sala,J., Yee,L., Malloy,M., Mishkin,M., and Courtney,S. (2004). Dissociable functional cortical topographies for working memory maintenance of voice identity and location. Cerebral Cortex *14*, 768-780.

Rapp,B., and Goldrick,M. (2000). Discreteness and interactivity in spoken word production. Psychological Review *107*, 460.

Rauschecker,J. (2011). An expanded role for the dorsal auditory pathway in sensorimotor control and integration. Hearing Research *271*, 16-25.

Rauschecker,J., and Tian,B. (2000). Mechanisms and streams for processing of 'What' and 'Where' in auditory cortex. Proceedings of the National Academy of Sciences *97*, 11800-11806.

Remez,R., Fellowes,J., and Rubin,P. (1997). Talker identification based on phonetic information. Journal of Experimental Psychology: Human Perception and Performance *23*, 651.

Rendall,D., Kollias,S., Ney,C., and Lloyd,P. (2005). Pitch (F0) and formant profiles of human vowels and vowel-like baboon grunts: the role of vocalizer body size and voice-acoustic allometry. The Journal of the Acoustical Society of America *117*, 944-955.

Rey,G., Levin,B., Rodas,R., Bowen,B., and Nedd,K. (1994). A longitudinal examination of crossed aphasia. Arch Neurol *51*, 95-00.

Rizzolatti,G., and Arbib,M. (1998). Language within our grasp. Trends in neurosciences *21*, 188-194.

Rizzolatti,G., and Craighero,L. (2004). The Mirror-Neuron System. Annu. Rev. Neurosci. *27*, 169-192.

Rizzolatti,G., Fadiga,L., Gallese,V., and Fogassi,L. (1996). Premotor cortex and the recognition of motor actions. Cognitive Brain Research *3*, 131-141.

Robey,R. (1998). A Meta-Analysis of Clinical Outcomes in the Treatment of Aphasia. J Speech Lang Hear Res *41*, 172-187.

Robinson,G., Rossor,M., and Cipolotti,L. (1999). Selective sparing of verb naming in a case of severe Alzheimer's disease. Cortex *35*, 443-450.

Roeltgen,D., and Heilman,K. (1984). Lexical agraphia: further support for the two-system hypothesus of linguistic agraphaia. Brain *107*, 811-827.

Rogalsky,C., Matchin,W., and Hickok,G. (2008). Broca's area, sentence comprehension, and working memory: an fMRI study. Frontiers in Human Neuroscience *2*.

Romanski,L. (2012). Integration of faces and vocalizations in ventral prefrontal cortex: implications for the evolution of audiovisual speech. Proceedings of the National Academy of Sciences *109*, 10717-10724.

Romanski,L., Bates,J., and Goldman-Rakic,P. (1999). Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. Journal of Comparative Neurology *403*, 141-157.

Rorden,C. MRIcron [Computer software]. 2007.

Ross,E. (1981). The aprosodias: Functional-anatomic organization of the affective components of language in the right hemisphere. Arch Neurol *38*, 561-569.

Ross,K., and Wertz,R. (2003). Quality of life with and without aphasia. Aphasiology *17*, 355-364.

Rousseeuw,P., and Croux,C. (1993). Alternatives to the median absolute deviation. Journal of the American Statistical Association *88*, 1273-1283.

Ryan,J., and Lopez,S. (2001). Wechsler Adult Intelligence Scale-III. In Understanding Psychological Assessment, Springer), pp. 19-42.

Saffran,J. (2001). The use of predictive dependencies in language learning. Journal of Memory and Language *44*, 493-515.

Sandt-Koenderman,W. (2011). Aphasia rehabilitation and the role of computer technology: Can we keep up with modern times? International journal of speech-language pathology *13*, 21-27.

Sarno,M. (1993). Aphasia rehabilitation: Psychosocial and ethical considerations. Aphasiology *7*, 321-334.

Sarno,M. (1997). Quality of life in aphasia in the first post-stroke year. Aphasiology *11*, 665-679.

Saur,D., Kreher,B., Schnell,S., Kummerer,D., Kellmeyer,P., Vry,M., Umarova,R., Musso,M., Glauche,V., and Abel,S. (2008). Ventral and dorsal pathways for language. Proceedings of the National Academy of Sciences *105*, 18035-18040.

Saur,D., Lange,R., Baumgaertner,A., Schraknepper,V., Willmes,K., Rijntjes,M., and Weiller,C. (2006). Dynamics of language reorganization after stroke. Brain *129*, 1371-1384.

Schall,S., Kiebel,S., Maess,B., and von Kriegstein,K. (2014). Voice Identity Recognition: Functional Division of the Right STS and Its Behavioral Relevance. Journal of Cognitive Neuroscience.

Schneider,W., Eschman,A., and Zuccolotto,A. E-Prime 2.0 software. 2012. Pittsburgh, PA, Psychology Software Tools.

Scott,S. (2008). Voice processing in monkey and human brains. Trends in Cognitive Sciences *12*, 323-325.

Scott,S., and Johnsrude,I. (2003). The neuroanatomical and functional organization of speech perception. Trends in neurosciences *26*, 100-107.

Scott,S., and Wise,R. (2003). PET and fMRI studies of the neural basis of speech perception. Speech Communication *41*, 23-34.

Scott,S., and Wise,R. (2004). The functional neuroanatomy of prelexical processing in speech perception. Cognition *92*, 13-45.

Seghier,M. (2013). The angular gyrus multiple functions and multiple subdivisions. The Neuroscientist *19*, 43-61.

Seghier,M., Ramlackhansingh,A., Crinion,J., Leff,A., and Price,C. (2008). Lesion identification using unified segmentation-normalisation models and fuzzy clustering. NeuroImage *41*, 1253-1266.

Sekiyama,K., Kanno,I., Miura,S., and Sugita,Y. (2003). Auditory-visual speech perception examined by fMRI and PET. Neuroscience Research *47*, 277-287.

Shallice,T. (1984). More functionally isolable subsystems but fewer modules? Cognition *17*, 243-252.

Shallice,T., and Vallar,G. (1990). The impairment of auditory-verbal short-term storage. Neuropsychological impairments of short-term memory 11-53.

Shaywitz,B., Shaywitz,S., Pugh,K., Mencl,W., Fulbright,R., Skudlarski,P., Constable,R., Marchione,K., Fletcher,J., and Lyon,G. (2002). Disruption of posterior brain systems for reading in children with developmental dyslexia. Biological psychiatry *52*, 101-110.

Shaywitz,S., Mody,M., and Shaywitz,B. (2006). Neural mechanisms in dyslexia. Current Directions in Psychological Science *15*, 278-281.

Shaywitz,S., Shaywitz,B., Fulbright,R., Skudlarski,P., Mencl,W., Constable,R., Pugh,K., Holahan,J., Marchione,K., and Fletcher,J. (2003). Neural systems for compensation and persistence: young adult outcome of childhood reading disability. Biological psychiatry *54*, 25-33.

Shaywitz,S., Shaywitz,B., Pugh,K., Fulbright,R., Constable,R., encl,W., hankweiler,D., iberman,A., kudlarski,P., and letcher,J. (1998). Functional disruption in the organization of the brain for reading in dyslexia. Proceedings of the National Academy of Sciences *95*, 2636-2641.

Shewan,C., and Kertesz,A. (1984). Effects of speech and language treatment on recovery from aphasia. Brain and Language *23*, 272-299.

Siegel,J., Power,J., Dubis,J., Vogel,A., Church,J., Schlaggar,B., and Petersen,S. (2014). Statistical improvements in functional magnetic resonance imaging analyses produced by censoring high-motion data points. Human Brain Mapping *35*, 1981-1996.

Simon,O., Mangin,J.-F., Cohen,L., Le Bihan,D., and Dehaene,S. (2002). Topographical layout of hand, eye, calculation, and language-related areas in the human parietal lobe. Neuron *33*, 475-487.

Sinanovic,O., Mrkonjic,Z., Zukic,S., Vidovic,M., and Imamovic,K. (2011). Post-stroke language disorders. Acta clinica Croatica *50*, 79-93.

Singh-Curry,V., and Husain,M. (2009). The functional role of the inferior parietal lobe in the dorsal and ventral stream dichotomy. Neuropsychologia *47*, 1434-1448.

Skeide,M., Brauer,J., and Friederici,A. (2014). Syntax gradually segregates from semantics in the developing brain. NeuroImage *100*, 106-111.

Smith,R. The role of fine phonetic detail in word segmentation. 2004. University of Cambridge.

Smith,S., and Brady,J. (1997). SUSAN - a new approach to low level image processing. International Journal of Computer Vision *23*, 45-78.

Speakability. Factsheets about aphasia and Speakablility. 2006.

Specht,K. (2014). Neuronal basis of speech comprehension. Hearing Research *307*, 121-135.

Staudt,M., Grodd,W., Niemann,G., Wildgruber,D., Erb,M., and Krageloh-Mann,I. (2001). Early left periventricular brain lesions induce right hemispheric organization of speech. Neurology *57*, 122-125.

Steele,R., Aftonomos,L., and Munk,M. (2003). Evaluation and treatment of aphasia among the elderly with stroke. Topics in Geriatric Rehabilitation *19*, 98-108.

Stefanis,L., Desmond,D., and Tatemichi,T. (1997). Crossed conduction aphasia associated with impairment of visuospatial memory. Neurocase *3*, 201-207.

Stevens,K., and Klatt,D. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. The Journal of the Acoustical Society of America *55*, 653-659.

Strand,E. Gender Stereotype Effects in Speech Processing. 2000. The Ohio State University.

Swanson,H., Harris,K., and Graham,S. (2013). Handbook of learning disabilities (London: Guilford Press).

Sweet,R., Dorph-Petersen,K., and Lewis,D. (2005). Mapping auditory core, lateral belt, and parabelt cortices in the human superior temporal gyrus. Journal of Comparative Neurology *491*, 270-289.

Swinburn,K., Porter,G., and Howard,D. (2004). CAT: Comprehensive Aphasia Test (UK: Psychology Press).

Talavage,T., Sereno,M., Melcher,J., Ledden,P., Rosen,B., and Dale,A. (2004). Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. Journal of neurophysiology *91*, 1282-1296.

Tanner,D. (2003). Eclectic perspectives on the psychology of aphasia. Journal of allied health *32*, 256-260.

Tanridag,O., and Kirshner,H. (1985). Aphasia and agraphia in lesions of the posterior internal capsule and putamen. Neurology *35*, 1797.

Teasell,R., Bitensky,J., Salter,K., and Bayona,N. (2005). The role of timing and intensity of rehabilitation therapies. Topics in stroke rehabilitation *12*, 46-57.

Titze,I. (1994). Principles of voice production (New Jersey: Prentice hall).

Townsend,N., Wickramasinghe,K., Bhatnagar,P., Smolina,K., Nichols,M., Leal,J., Luengo-Fernandez,R., and Rayer,M. (2012). Coronary heart disease statistics: 2012 edition (London: British Heart Foundation), p. 57.

Turkeltaub,P., and Coslett,H. (2010). Localization of sublexical speech perception components. Brain and Language *114*, 1-15.

Turkeltaub,P., Coslett,H., Thomas,A., Faseyitan,O., Benson,J., Norise,C., and Hamilton,R. (2012). The right hemisphere is not unitary in its role in aphasia recovery. Cortex *48*, 1179-1186.

Turner-Stokes,L., Kalmus,M., Hirani,D., and Clegg,F. (2005). The Depression Intensity Scale Circles (DISCs): a first evaluation of a simple assessment tool for depression in the context of brain injury. Journal of Neurology, Neurosurgery & Psychiatry *76*, 1273-1278.

van der Gaag,A., Smith,L., Davis,S., Moss,B., Cornelius,V., Laing,S., and Mowles,C. (2005). Therapy and support services for people with long-term stroke and aphasia and their relatives: a six-month follow-up study. Clinical Rehabilitation *19*, 372-380.

Van Lancker,D., and Canter,G. (1982). Impairment of voice and face recognition in patients with hemispheric damage. Brain and Cognition *1*, 185-195.

Van Lancker,D., Cummings,J., Kreiman,J., and Dobkin,B. (1988). Phonagnosia: a dissociation between familiar and unfamiliar voices. Cortex *24*, 195-209.

Van Lancker,D., Kreiman,J., and Cummings,J. (1989). Voice perception deficits: Neuroanatomical correlates of phonagnosia. Journal of clinical and experimental neuropsychology *11*, 665-674.

Van Lancker,D., and Sims,J. (1992). The identifiation of affective-prosodic stimuli by left-and righthemisphere-damaged subjects: All errors are not created equal. Journal of Speech and Hearing Research *35*, 963-970.

Vigneau,M., Beaucousin,V., Herve,P., Duffau,H., Crivello,F., Houde,O., Mazoyer,B., and Tzourio-Mazoyer,N. (2006). Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. NeuroImage *30*, 1414-1432.

Vogt,B., Berger,G., and Derbyshire,S. (2003). Structural and functional dichotomy of human midcingulate cortex. European journal of neuroscience *18*, 3134-3144.

von Kriegstein,K., Eger,E., Kleinschmidt,A., and Giraud,A. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. Cognitive Brain Research *17*, 48-55.

von Kriegstein,K., and Giraud,A. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. NeuroImage *22*, 948-955.

Wada,J., and Rasmussen,T. (1960). Intracarotid injection of sodium amytal for the lateralization of cerebral speech dominance: experimental and clinical observations. Journal of Neurosurgery *17*, 266-282.

Wade,D., Hewer,R., David,R., and Enderby,P. (1986). Aphasia after stroke: natural history and associated deficits. Journal of Neurology, Neurosurgery & Psychiatry *49*, 11-16.

Wade,T., and Holt,L. (2005). Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task. The Journal of the Acoustical Society of America *118*, 2618-2633.

Walker,S., Bruce,V., and O'Malley,C. (1995). Facial identity and facial speech processing: Familiar faces and voices in the McGurk effect. Perception & Psychophysics *57*, 1124-1133.

Warren,J., Crinion,J., Lambon R., Matthew,B., and Wise,R. (2009). Anterior temporal lobe connectivity correlates with functional outcome after aphasic stroke. Brain *132*, 3428-3442.

Warren,J., Uppenkamp,S., Patterson,R., and Griffiths,T. (2003). Separating pitch chroma and pitch height in the human brain. Proceedings of the National Academy of Sciences *100*, 10038-10042.

Warren,R. (1970). Perceptual restoration of missing speech sounds. Science *167*, 392-393.

Warrier,C., Wong,P., Penhune,V., Zatorre,R., Parrish,T., Abrams,D., and Kraus,N. (2009). Relating structure to function: Heschl's gyrus and acoustic processing. The Journal of Neuroscience *29*, 61-69.

Watson,R., Latinus,M., Noguchi,T., Garrod,O., Crabbe,F., and Belin,P. (2014). Crossmodal adaptation in right posterior superior temporal sulcus during face-voice emotional integration. The Journal of Neuroscience *34*, 6813-6821.

Wechsler,D. (2008). Wechsler Adult Intelligence Scale-IV (New York: Psychological Corporation).

Wechsler,D. (1999). Wechsler Memory Scale -3rd Edition (WMS-III) (New York: Psychological Corporation).

Weisenberg,T., and McBride,K. (1935). Aphasia: A Clinical and Psychological Study (New York: The Commonwealth Fund).

Wernicke,C. (1874). The symptom complex of aphasia: A psychological study on an anatomical basis. Arch Neurol *22*, 280.

Whiteley,W., Lindley,R., Wardlaw,J., and Sandercock,P. (2006). Third international stroke trial. international Journal of Stroke *1*, 172-176.

Whiteside,S. (1998). The identification of a speaker's sex from synthesized vowels. Perceptual and motor skills *87*, 595-600.

Whitworth,A., Webster,J., and Howard,D. (2014). A Cognitive Neuropsychological Approach to Assessment and Intervention in Aphasia: A Clinician's Guide (East Sussex: Psychology Press).

Wilcox,R. (2012). Introduction to robust estimation and hypothesis testing (Oxford: Academic Press).

Wildgruber,D., Ackermann,H., Klein,M., Riecker,A., and Grodd,W. (2000). Brain activation during identification of affective speech melody: influence of emotional valence and sex. NeuroImage *11*, S341.

Wile,D., and Balaban,E. (2007). An auditory neural correlate suggests a mechanism underlying holistic pitch perception. PLoS ONE *2*, e369.

Wilke,M., and Lidzba,K. (2007). LI-tool: a new toolbox to assess lateralization in functional MR-data. Journal of neuroscience methods *163*, 128-136.

Willmes,K., and Poeck,K. (1993). To what extent can aphasic syndromes be localized? Brain *116*, 1527-1540.

Wilson,B., Alderman,N., Burgess,P., Emslie,H., and Evans,J. (1996). Behavioural Assessment of the Dysexecutive Syndrome (Bury St Edmunds: Thames Valley Test Company).

Wise,R., Chollet,F., Hadar,U., Friston,K., Hoffner,E., and Frackowiak,R. (1991). Distribution of cortical neural networks involved in word comprehension and word retrieval. Brain *114*, 1803-1817.

Wise,R., Scott,S., Blank,S., Mummery,C., Murphy,K., and Warburton,E. (2001). Separate neural subsystems within `Wernicke's area'. Brain *124*, 83-95.

Zatorre,R., Belin,P., and Penhune,V. (2002). Structure and function of auditory cortex: music and speech. Trends in Cognitive Sciences *6*, 37-46.

Zatorre,R., Meyer,E., Gjedde,A., and Evans,A. (1996). PET Studies of Phonetic Processing of Speech: Review, Replication, and Reanalysis. Cerebral Cortex *6*, 21-30.