Rehabilitation in aphasia: Using brain and behavioural measures to investigate the effects of a digital speech comprehension therapy

Victoria Ann Alexandra Fleming

2019

UCL Queen Square Institute of Neurology
University College London

This thesis is submitted for the degree of Doctor of Philosophy

Supervisors:
Professor Alex Leff
Dr Peter Zeidman

Declaration of Authorship

I, Victoria Fleming, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgements

I am thankful to UCL for providing me with a studentship to complete my PhD at the Queen Square Institute of Neurology.

My first thanks goes to Professor Alex Leff, who gave me the opportunity to join the Neurotherapeutics Lab as a research assistant, and encouraged me to think about a completing a PhD. I am grateful for receiving this opportunity, and for your continued encouragement, support, and advice throughout the ups and downs of this project. I could not have completed this PhD without your generous feedback and suggestions.

I am also grateful to Dr Sonia Brownsett for giving me invaluable insight into the research world. Your curiosity into aphasia, and genuine compassion for those we are trying to help, encouraged me to embark on this PhD journey. I missed your passion and drive in the last months of this project, and hope that our paths will cross again in the research world in the future.

My appreciation also extends to the many people with aphasia who have been so generous with their time and interest in this project. Every single individual gave valuable input along the way, and made this project possible. Their input will undoubtedly help others in the future.

Thank-you also to everyone in the Neurotherapeutics Lab. Over the last four years I have been lucky to receive generous advice, friendship and support for which I am grateful, Anna, Zoe, Davide, Sheila, Sharifa, Katerina, Oscar, Jenny, Henry, Maria, David, Yean, Emily, Tom, and Catherine.

For help in my data analyses I also owe thanks to Dr Oscar Aguilar, for patiently helping me whilst at the end stages of his own thesis (I know how that feels now!), Davide Nardo for his assistance with many scripts, Professor John Ashburner for advice on longitudinal imaging, and Dr Tom Hope for some last-minute suggestions. I also owe thanks to the PLORAS team for collaborating on this project: Professor Cathy Price, Zula, Sophie, Hayley, Kate, and Shumi.

My final thanks goes to my family for their continued encouragement. And Ed and Ginger, for their unwavering patience and support.

Abstract

This thesis investigates the development and testing of a novel tablet-based speech comprehension therapy, for persons with chronic aphasia (PWA).

In Chapter 1 of this thesis, I present a qualitative study on the development of a speech comprehension therapy app (Listen-In), with gamification. Five co-design focus groups were carried out, resulting in an iterative cycle of prototype development. Using thematic analysis, a number of barriers to usability and enjoyability were identified. This resulted in a multitude of design changes, which led to a final product suitable to be self-administered by patients.

In Chapter 2, I analysed data from a cross-over clinical trial of Listen-In (N=35), which compared 12-weeks of Listen-In treatment with 12-weeks of standard care. These findings showed: (i) PWA made large and significant gains in speech comprehension for words trained during treatment; (ii) gains were item specific, suggesting facilitation of item-specific networks; (iii) a combination of baseline measures explained only a small amount of variability in treatment outcomes.

In Chapter 3, I used voxel-based morphometry (VBM) to investigate whether baseline structural integrity contributed to treatment outcomes, in a subgroup of patients (N=25). I found that greater volume of mainly white matter, in distributed regions in the right hemisphere, predicted greater response to treatment. I speculated that these regions related to multi-functional networks supporting a range of cognitive and language functions.

In Chapter 4, I used longitudinal VBM to investigate therapy-driven structural neuroplasticity in the same subgroup of patients. I found that greater improvements in comprehension of treated items, were related to tissue changes in bilateral temporal lobes, in key speech processing regions.

In the final discussion section, I interpret these key findings, their clinical implications, and directions for future research.

Impact Statement

There are over 1.2 million stroke survivors in the UK, and around a third of people who have a stroke experience aphasia, a language impairment which can have a considerable impact on all aspects of language and communication (*State of the Nation, Stroke Statistics*, 2017). For some, language skills improve following a stroke; however, for many, these impairments remain chronic. The evidence shows that speech and language treatments can have a significant and positive impact on language and communication skills in persons with aphasia (Brady, Kelly, Godwin, Enderby, & Campbell, 2016), and that high dose treatments, of approximately 100 hours, confer the greatest benefits (Bhogal, Teasell, & Speechley, 2003). However, due to limited healthcare resources, a typical patient in the UK receives less than 10 hours of speech and language therapy (Code & Heron, 2003). Furthermore, few studies have investigated therapies which specifically target speech comprehension, meaning the efficacy of this type of treatment is unknown.

The first part of my thesis investigated the development and testing of a novel speech comprehension therapy app ('Listen-In), for persons with aphasia. This resulted in a final product which was co-designed with persons with aphasia, and the results of the clinical trial showed that patients made significant improvements in comprehension of words they trained on during treatment. Listen-In will now be made available to the public. As such, the development of this application provides clear impact to the wider national and international healthcare and patient community by:

- Delivering an app which has been specifically co-designed with and for patients;
- Providing public access to the first evidence-based app specifically targeting speech comprehension;
- Enabling patients to self-administer treatment and achieve the high recommended dose;
- Providing clinicians with a novel therapeutic tool which is feasible to use within current service limitations.

The findings in this thesis also provide a significant and novel contribution to the academic field of aphasia and its rehabilitation, by showing that speech comprehension can be significantly improved in persons with aphasia, even years after stroke. Previously, few studies have investigated the efficacy of this type of treatment, therefore these findings open up a new possible avenue of therapeutic aphasia research. A novel finding in this research was evidence of structural neuroplasticity, driven by therapy. This discovery shows that specific and targeted treatment can support mechanisms of neuroplasticity in patients with long-standing aphasia. These findings open up new possibilities for research in neuroscience-based therapeutic interventions. The longitudinal method in this study has not been frequently reported in patient studies (Ashburner, 2013). The positive findings show this may be a sensitive tool for future investigations of structural plasticity in response to treatment, both in aphasic patients, and other populations. These findings will continue to impact the research community through dissemination at national and international conferences and research journals, and will provide significant scope for future research collaborations with technology partners and patient communities.

Table of Contents

Statement of contribution

The Listen-In project was funded by the National Institute for Health Research, Invention for Innovation (i4i) program. Professor Alex Leff and Dr Sonia Brownsett developed the grant for this project, which was carried out between 2015 and 2018. I conducted the clinical trial alongside Professor Alex Leff and Sonia Brownsett. I collaborated with Dr Holly Robson to develop the Phoneme Discrimination Test (PDT). Yean-Hoon Ong programmed the therapy component of Listen-In, and the PDT. Davide Nardo programmed the Auditory Comprehension Test (ACT). Maria Andree Maegli Quezada assisted in developing the ACT and Henry Coley-Fisher assisted in developing the Environmental Sounds Test. SOFT-V were financially reimbursed for their services in developing the Listen-In app. The raw data collected from this project forms the basis of my thesis. All data pre-processing, statistical analyses, and interpretations are my own, and all other contributions are appropriately referenced.

List of figures

List of tables

List of abbreviations

A1	Auditory cortex
ANOVA	Analysis of variance
ACC	Anterior cingulate cortex
ALI	Automatic lesion identification
ATL	Anterior temporal lobe
ACT	Auditory Comprehension Test
ALM	Automatic linear modelling
BPVS	British Picture Vocabulary Scale
CAT	Comprehensive Aphasia Test
CIAT	Constraint induced aphasia therapy
CC	Corpus callosum
CATTELL	Culture Fair Test of Intelligence
DTI	Diffusion tensor imaging
DLPFC	Dorsolateral prefrontal cortex
ENVASA	Environmental Sounds Test
FWE	Family wise error
FG	Focus group
FA	Fractional anisotropy
fMRI	Functional magnetic resonance imaging
G1	Group 1
G2	Group 2
GM	Grey matter
IFG	Inferior frontal gyrus
IFOF	Inferior fronto-occipital fasciculous
ILF	Inferior longitudinal fasciculous
IPL	Inferior parietal lobe
ITG	Inferior temporal gyrus
LH	Left hemisphere
MR	Magnetic resonance

MRI	Magnetic resonance imaging
M	Mean
MIT	Melodic Intonation Therapy
MRIcron	MRI cross-platform NIfTI format image viewer
MTG	Middle temporal gyrus
MNI	Montreal Neurological Institute
N	Number of participants
P1	Participant 1
PWA	Persons with aphasia
PDT	Phoneme Discrimination Test
pSTG	Posterior superior temporal gyrus
pSTS	Posterior superior temporal sulcus
Pre-SMA	Pre-supplementary motor area
RH	Right hemisphere
SD	Standard deviation
SPSS	Statistical package for the social sciences
SPM12	Statistical and Parametric Mapping Software 12
STG	Superior temporal gyrus
STS	Superior temporal sulcus
SMA	Supplementary motor area
SART	Sustained Attention to Response Task
T1	Timepoint 1
VLPFC	Ventrolateral prefrontal cortex
VBM	Voxel based morphometry
VBQ	Voxel based quantification
WM	White matter

Glossary

Lexical item	A single word e.g. cat
Challenge	A particular spoken stimuli-to-picture trial in Listen-In therapy or the ACT, used to train a lexical item
Linguistic construction	Phrase or sentence used to train a lexical item in Listen-In and the ACT
Trained Challenge Item (Trained items)	110 lexical items tested in the ACT, and trained in Listen-In in the same linguistic construction
Trained Lexical Item	A lexical item tested in the ACT, but trained in Listen-In in a different linguistic construction
Untrained items	110 items not trained in Listen-In, but tested in the ACT
Core Therapy Challenge	Core therapy items trained during Listen-In treatment, but not tested in the ACT
Set A / Set B	110 trained / 110 untrained items Or 110 untrained / 110 trained items (half of patients randomised to each condition)

1 Introduction

The incidence of people suffering a stroke in the UK is >100,000 pa (*State of the Nation, Stroke Statistics*, 2017). Improving hyperacute and acute medical services mean that more and more people are surviving their stroke, so the prevalence continues to rise, with current estimates at 1.2 million stroke survivors in the UK (*State of the Nation, Stroke Statistics*, 2017). Around a third of people who have a stroke experience aphasia, a generalised language disorder which can affect speaking, understanding, reading and writing.

Speech comprehension is crucial to being able to communicate effectively, and can be a particularly devastating consequence of aphasia. Reduced ability to communicate can lead to reduced social networks and loss of friendships (Northcott & Hilari, 2011), reduced participation at home, at work, and in the community (Pike, Kritzinger, & Pillay, 2017), and reduced quality of life (Hilari, Wiggins, Roy, Byng, & Smith, 2003; Ross & Wertz, 2003). Impaired speech comprehension can also affect individuals' participation in other therapies (e.g. occupational and physiotherapy (Jette, Warren, & Wirtalla, 2005)) as well as participation in research (Cruice, Worrall, Hickson, Hirsch, & Holland, 2013).

Despite this, speech comprehension impairments and their rehabilitation have been relatively understudied in persons with aphasia (PWA). This is complicated by the finding that impairments in this area are often under-identified by family and friends, health professionals, and PWA themselves (Chwat & Gurland, 1981; Dorze & Brassard, 1995; McClenahan, Johnston, & Densham, 1992). One reason for this may be the relatively hidden nature of comprehension impairments compared to more obvious speech production deficits. For this reason, prevalence can be hard to estimate; however, one study found 48% of 122 acute stroke inpatients had impairments in spoken word comprehension (Breese & Hillis, 2004), and PWA can continue to experience difficulties understanding speech at one-year post stroke (Pedersen, Vinter, & Olsen, 2004).

Speech and language therapy (SLT) benefits people with aphasia across multiple areas of language, including receptive language (Brady et al., 2016). However, more detailed conclusions remain elusive due to insufficient evidence, particularly for speech comprehension therapies. Two approaches of significant clinical interest are intensity of therapy, and method of delivery. A review of 10 studies revealed that high dose SLT (around 100 hours) was beneficial over conventional SLT (Bhogal et al., 2003). Delivering 100 hours of therapy is not feasible for speech and language therapists, considering a typical patient receives, on average, just 12 hours of therapy in the UK (around 4 as an in-patient (Kavanagh, Stanley, & Tyrrell, 2017), and around 8 as an out-patient (Code & Heron, 2003)). Additionally, high intensity research studies face significantly higher drop-out rates (Brady et al., 2016). One reason for this could be the dull and repetitive nature of mass practice tasks in impairment-based therapies (Kurland, Wilkins, & Stokes, 2014; Varley et al., 2016).

Computer mediated delivery of SLT is a potential solution to the problems of under-dosing, and mundane nature of tasks. Reviews of the preliminary evidence base for technology based interventions have provided some promising findings in terms of treatment outcomes (Lavoie, Macoir, & Bier, 2017; Zheng, Lynch, & Taylor, 2015). However, the scientific evidence is slim, and it is still a relatively understudied approach (Brady et al., 2016). This is despite there being a wealth of computer programs and applications available (www.aphasiasoftwarefinder.org). Furthermore, many existing applications focus on speech production, and to the authors knowledge, no evidence-based apps specifically target speech comprehension impairments as their primary focus. Most existing therapy programs have mapped traditional tasks into a digital format, and few have utilised gamification techniques found in consumer gaming apps (e.g. Lumosity brain training, www.lumosity.com). One study investigated the use of virtual reality environment, EVA Park, which was co-created with aphasic patients. This was developed as a platform for social interaction and delivery of therapy (with a clinician) (Galliers et al., 2017). Participants demonstrated functional language gains after language stimulation sessions and also excellent compliance (Marshall et al., 2016). This suggests that use of novel software is feasible for this population, and may help improve user experience and

engagement; although in this case, as therapists are spending one-to-one time with patients, it does not solve the lack of therapist resource issues alluded to above.

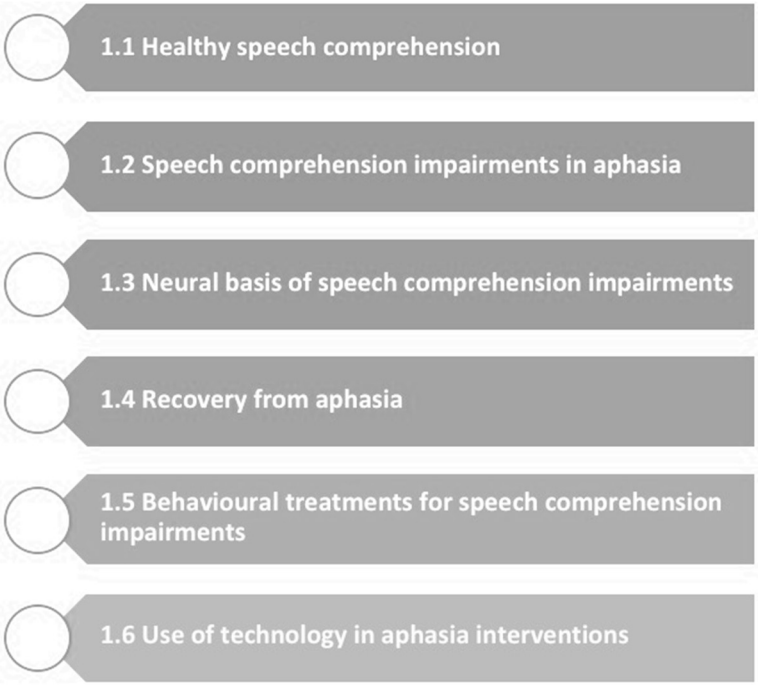
Developments in neuroimaging over the last few decades have revealed a wealth of information about the neural architecture of the language comprehension network in healthy individuals (Hickok, 2009; Price, 1998; Rauschecker & Scott, 2009; Scott, 2012; Scott, Blank, Rosen, & Wise, 2000; Turken & Dronkers, 2011), and investigations into the relation between lesion damage and language impairments has added considerably to our understanding of the neural substrates of speech comprehension (Dronkers, Wilkins, Valin, Redfern, & Jaeger, 2004; Wilson, 2016). However, there have been a relative lack of studies which have investigated the neural correlates of speech comprehension treatment. This is despite increasing interest in the neural correlates of naming treatments in aphasic patients (Abel, Weiller, Huber, Willmes, & Specht, 2015; McKinnon et al., 2017; Menke et al., 2009; Nardo, Holland, Leff, Price, & Crinion, 2017). One difficulty is that speech comprehension tasks require an extra overt response from patients to indicate the meaning of a word or sentence has been correctly understood; integrating these types of tasks within a scanner environment is often a challenge. One solution is to investigate long term structural changes which don't rely on online activities; however, longitudinal structural imaging studies are rare both within the aphasia population, and healthy individuals acquiring new skills. Nevertheless, several studies have recently reported changes in grey matter (Draganski et al., 2004) and white matter (Schlegel, Rudelson, & Tse, 2012) in response to learning new skills, including language, therefore one unexplored avenue in aphasia treatments is to investigate the possible structural neural correlates of therapy driven recovery. If identified, this line of enquiry may aid in increasing understanding of recovery mechanisms, and developing and refining neuroscience-based interventions.

Over the course of my PhD, I will contribute new evidence to the topics previously described by:

(1) Developing and testing a tablet-based speech comprehension therapy app, with gamification, to maximise practice time and thus, hopefully, clinical efficacy;

(2) Investigating the relation between brain structure and response to treatment in a group of individuals with chronic aphasia.

In the remainder of this introduction, this thesis will provide a background into the following areas:



1.1 Healthy Speech Comprehension

1.1.1 The auditory pathway: ear to cortex

In humans, speech perception begins at the ear. Air vibrations arrive at the outer ear, and travel to the tympanic membrane ('ear drum'). In the middle ear, three small bones (ossicles) transmit these vibrations to the oval window of the inner ear. From here, pressure waves continue in the fluid filled cochlea, where receptors transduce the mechanical waves into electrical signals. Fibres from both cochlea synapse in the brainstem at the cochlea nuclei, and then make contralateral and ipsilateral connections with the superior olivary complex. Fibres than travel via the inferior colliculus in the midbrain to the medial geniculate in the thalamus, and then to the auditory cortex (Recanzone & Sutter, 2011). The auditory system must then process the properties of the acoustic signal, to determine the location of the sound, its spectral and temporal properties, and what the sounds represent (Recanzone & Sutter, 2011).

1.1.2 The auditory cortex

Studies investigating the neural basis of audition have drawn on primate models, due to both the similarity of the primate and human auditory cortices, and similarity of the spectrotemporal structure of human and primate vocalisations (Rauschecker & Tian, 2000). In primates, the auditory cortex has been divided into core, belt, and parabelt regions, characterised by complex interconnections. This organisation is believed to reflect hierarchical processing, from simple receptive-field organisation, to complex perceptual processing (Rauschecker & Scott, 2009). The core, also divided into three cortical areas, contains the primary auditory cortex (A1), rostral field, and rostrotemporal field, organised tonotopically. Anatomically, A1 is located in the superior temporal plane. Information moves from A1 to adjacent belt, and then parabelt, fields. Neurons in belt fields show specialisation to different aspects of the signal, whilst parabelt regions are believed to process more complex acoustic features (Recanzone & Sutter, 2011). From here, there is a convergence of auditory information with other sensory domains via widely distributed pathways in the parietal and frontal lobes. Studies of the human

auditory cortex have demonstrated a broadly similar pattern of tonotopic organisation (Recanzone & Sutter, 2011). Hierarchical organisation occurs in parallel streams in relation to different sounds features. There is also evidence, in animal and humans, for short term plasticity in the auditory cortex, driven by both bottom up and top down inputs (Jääskeläinen, Ahveninen, Belliveau, Raij, & Sams, 2007) enabling auditory sensory memory representations, important for speech comprehension and working memory tasks (Jääskeläinen et al., 2007).

1.1.3 Dual processing streams

A dual stream account of auditory processing was first proposed following electrophysical and anatomical tract tracing studies in monkeys, which show neuronal response selectivities to species specific vocalisations, and their spatial locations (Rauschecker & Tian, 2000; Romanski et al., 1999; Tian et al., 2001). Two functionally and anatomically distinct processing streams were proposed in primates, akin to dorsal and ventral streams which have been well described within the visual system (Kaas & Hackett, 1999). A ventral 'what' pathway is involved in identifying auditory objects or complex patterns, mapping sound to meaning; whilst a dorsal 'where' pathway is involved in spatial location of sounds (Rauschecker & Tian, 2000; Romanski et al., 1999).

In humans, dual-stream frameworks have subsequently been put forward which posit two distinct auditory processing streams, both originating from the primary auditory cortex in Heschl's gyrus (HG) (Hickok & Poeppel, 2004, 2007; Rauschecker & Scott, 2009; Saur et al., 2008). More recently, evidence suggests that separation of the acoustic components of speech in the brainstem is a precursor to 'what' and 'where' cortical pathways (Jääskeläinen et al., 2007). These models have been developed and refined, in large part, due to the emergence of functional neuroimaging techniques, which have enabled investigators to localise brain regions specialised for different aspects of auditory input, including those specialised for speech. For example, in an early study, Mummery, Ashburner, Scott, & Wise (1999), using positron emission technology (PET), used a passive listening task in which six healthy participants listened to speech (words), versus noise with similar acoustic properties to speech, building on previous work which contrasted

speech with rest, or other types of noise. For both types of stimuli there was activation in the transverse temporal gyri and dorsal STG; but for speech, there was a larger region which encompassed the ventral STG/STS bilaterally (anterior and middle regions), with extension into the posterior STG/STS in the left hemisphere. These findings therefore distinguished regions which responded to both types of input, as well as regions which were specialised for speech. Since then, a wealth of studies have been performed with different types of contrasts, and it is generally accepted that speech perception is subserved by two main processing pathways. However, there is still considerable debate on hemispheric contributions to speech processing,

1.1.3.1 Ventral stream

The ventral stream has been proposed to map sound onto meaning. It is bilaterally organised, and projects to middle and inferior temporal regions, and inferior frontal regions (Rauschecker & Scott, 2009). Evidence for a ventral route in humans has emerged from functional imaging studies which have tasked participants with listening a range of auditory stimuli and contrasted brain activations between these conditions. Typically, tasks have involved contrasts with meaningful speech. In an illustrative study, Scott and colleagues (2000) utilised PET imaging during a passive listening task. Four different contrasts represented intelligible stimuli (speech, noise-vocoded speech) and non-intelligible stimuli which still had similar acoustic properties to speech (rotated speech, and rotated noise-vocoded speech). The findings revealed different activation patterns: the pSTS and ventrolateral STG were equally activated by stimuli containing phonetic information; whereas the anterior left STS was preferentially activated by intelligible speech, reflecting a ventral "what" pathway. These results demonstrate a left-dominant response for intelligible stimuli in anterior temporal regions, supported by further studies (Evans et al., 2014; Narain et al., 2003).

In a meta-analysis of over 100 functional imaging studies, DeWitt and Rauschecker (2012) propose that phoneme processing is associated with the left mid-STG, whilst the integration of phonemes into whole words is localised to the left anterior STG: "The segregation of phoneme and word-form processing along STG implies a growing encoding

specificity for complex phonetic forms by higher-order ventral-stream areas" (p.E511, DeWitt & Rauschecker, 2012). These findings support others who have suggested an antero-lateral gradient of speech processing along the ventral route, as auditory stimuli become increasingly more complex, from simple tones to speech (Binder et al., 2000; Mummery et al., 1999; Scott et al., 2000). More specifically, phoneme processing refers to brain regions which are sensitive to the low-level spectrotemporal properties of speech sounds, and has been proposed to rely on neurons within the mid-STG which have receptive-field tuning for phonetic speech sounds (DeWitt & Rauschecker, 2012). Word-form recognition refers to sensitivity to the temporal arrangements of phonemes and therefore higher-order features of speech, and has been localised to the anterior STG/STS (DeWitt & Rauschecker, 2012).

Following word-form recognition, auditory information is thought to converge with more abstract semantic information, enabling extraction of word meanings (Patterson, Nestor, & Rogers, 2007). In a meta-analysis of functional imaging studies, word and non-word contrasts showed brain activation patterns in regions adjacent to, but distinct from, auditory regions, in the left angular gyrus, left middle and inferior temporal gyri, and bilateral temporal poles (Binder et al., 2000). These findings are in line with bilateral atrophy of the temporal lobes seen in semantic dementia patients (Mion et al., 2010), and others who have demonstrated a key role of the anterior temporal lobes in conceptual knowledge (Lambon Ralph, Pobric, & Jefferies, 2009; Robson et al., 2014). However, others have suggested that lexical semantic access, in the form of a "sound-to-meaning" interface, is localised in posterior lateral temporal lobes (specifically in the region of the pMTG), mainly in the LH but with some functions also likely to be held in the RH (Hickok & Poeppel, 2007). The previous descriptions of ventral stream processing have assumed a serial organisation; however, at a neural level, there is insufficient evidence for the neural computations which are carried out, and whether these occur serially, or in parallel streams (Hickok & Poeppel, 2007).

The localisation of speech processing to mid and anterior temporal regions contrasts with the traditional view that Wernicke's area, in the region of the pSTG, underlies speech comprehension. It has been argued that the accumulation of new evidence from more

tightly controlled studies with more advanced neuroimaging techniques supports this conclusion (DeWitt & Rauschecker, 2012). However, others have argued that many studies have relied on sentence level stimuli and low-level auditory contrasts, and cite evidence which show that lesions to posterior STG/STS lead to auditory comprehension deficits (Hickok & Poeppel, 2007). The role of the posterior STG/STS within dual stream models is therefore a matter of debate.

1.1.3.1 Dorsal stream

The dorsal stream has been proposed to map sound onto articulatory representations. It projects from the superior temporal plane to the inferior partial cortex, and inferior frontal cortex (Rauschecker & Scott, 2009). In primates, this pathway has been defined as a "where" pathway, enabling spatial localisation of sounds. In humans, a number of functional imaging studies also support a key role of this stream in sound localisation and motion-in-space (for a review see Rauschecker (2011)). In humans, it has been proposed to subserve auditory-motor integration for speech, strongly lateralised to the left hemisphere (Hickok & Poeppel, 2000, 2004, 2007; Rauschecker & Scott, 2009; Wise et al., 2001). This network includes the STS, left IFG, and a region at the temporoparietal junction; the latter region has been proposed to be a sensory-motor interface between the posterior temporal cortex and motor speech representations (Hickok, 2009; Wise et al., 2001). Given the dual functions of the dorsal stream in spatial localisation and sensori-motor integration, a further proposal is that the planum temporale operates as a 'computational hub', processing many types of complex sounds, including speech (Griffiths & Warren, 2002). This is in contrast to Hickok and Poeppel (2007) who describe the planum temporal in terms of sensorimotor integration relating to the vocal tract, rather than auditory-motor integration.

1.1.3.1 Neural architecture of dual stream pathways

The arcuate fasciculous and lateral superior longitudinal fasciculous have been pinpointed as underlying the dorsal stream, connecting Broca's area with posterior temporal regions (Catani, Jones, & Ffytche, 2005; Saur et al., 2008). These white matter pathways are

therefore believed to support phonological and articulatory processes. Conversely, a number of different fibre pathways have been linked to the ventral stream, and as yet, the precise function of how these fibres functionally relate to language processing is unclear. These fibre pathways include the uncinate fasciculous, inferior longitudinal fasciculous, middle longitudinal fasciculous, and the extreme capsule (Rauschecker & Scott, 2009; Saur et al., 2008). More recently, the inferior fronto-occipital fasciculous has been identified in humans, connecting the frontal lobe with occipital and parietal lobes. This pathway has been postulated to underlie semantic processing due to its connectivity with cortical regions implicated in semantic processes (Vigneau et al., 2006). It has also been suggested as the main anatomical substrate of the ventral language pathway (Almairac, Herbet, Moritz-Gasser, de Champfleur, & Duffau, 2015).

1.1.3.1 Hemispheric contributions

Previously, a number of studies have shown activations for speech which are confined to the LH, leading to hypotheses that speech processing is lateralised to mainly the LH (e.g. (Scott et al., 2000). However, other studies have demonstrated bilateral activations. For example, in a key study, Okada and colleagues (2010) replicated Scott and colleagues (2000) design, but used a method which was more sensitive to within-subject activity (multivariate pattern analysis). The authors found bilateral activity, downstream from auditory regions, in both the anterior STS, as well as the posterior STS, which was sensitive to speech intelligibility, and less sensitive to acoustic variation. These findings support bilateral and posterior processing by demonstrating speech recognition along the length of the STS in both hemispheres.

These findings have led some researchers have put forward models which emphasise bilateral speech recognition (Hickok & Poeppel, 2000, 2004, 2007). Within this model, the dorsal stream is strongly left dominant, whilst acoustic-phonetic processing along the ventral stream is proposed to occur in bilateral and parallel pathways which map sound onto meaning in both hemispheres. Hickok and Poeppel (2007) argue that this accounts for the finding that unilateral lesions do not usually lead to severe speech recognition difficulties. For example, patients with left hemisphere lesions and auditory

comprehension deficits rarely show completely impaired word comprehension ability, and have been reported to identify between 70-80% of single words in matching tasks with a 25% chance of success (Hickok et al., 2008). In a novel study, Hickok and colleagues (2008) investigated the contribution of both hemispheres to speech comprehension in healthy individuals, by administering sodium amytal injections (anaesthetic) to each hemisphere independently. Participants made significantly more errors during LH injection, and these errors were mainly semantic (75%, like aphasic patients) rather than phonological. The authors interpret this finding as evidence supporting bilateral speech sound processing in healthy individuals. Bilateral speech processing is also supported by a wealth of studies, in healthy individuals, which demonstrate right hemisphere activation patterns during language tasks, but which are mainly bilateral in nature (Vigneau et al., 2011). Turken and Dronkers (2011) also demonstrate that speech comprehension regions in the left hemisphere show both structural and functional connectivity with right hemisphere homologues, supporting the presence of interhemispheric connectivity during speech comprehension. The roles of both hemispheres, however, may be computationally different, so that each carries out different aspects of speech processing, allowing for redundancy of phonological cues (Hickok et al., 2008; Hickok & Poeppel, 2007). Left-dominance emerges during later stages where the network interfaces with widely distributed semantic representations, and this post-phonemic level of breakdown is postulated to be the cause of speech comprehension impairments. However, this model is not without its critics, and others have found evidence for deficits in acoustic-phonological processing in Wernicke's aphasia (WA) patients, which correlate with speech comprehension ability (Robson, Davies, Ralph, & Sage, 2012; Robson, Grube, Lambon Ralph, Griffiths, & Sage, 2013; Robson, Keidel, Lambon Ralph, & Sage, 2012).

1.2 Speech Comprehension Impairments in Aphasia

The scientific study of language is widely claimed to have begun by Pierre Paul Broca in the 1860s (Broca, 1861a, 1861b, 1864). His series of case studies demonstrated that loss of speech production was associated with a specific part of the left frontal lobe (posterior inferior frontal gyrus (IFG)). Around the same time, Carl Wernicke associated an area of the left posterior temporal gyrus with speech comprehension disorders. Incorporating

Broca's findings, he went on to publish 'Der Aphasische Symptomencomplex' in 1874 (Wernicke, 1908) the first theoretical framework which classified different kinds of aphasia. Today, although there is no universally accepted classification system, these broad classifications are widely adopted (I.E. Broca's, Wernicke's, global and conduction aphasia, amongst others); however, most experts in the field accept the limitations of this historical classification system.

1.2.1 Cognitive neuropsychological model of speech comprehension impairments

Cognitive neuropsychology developed later on in the 1970s, beginning with studies of patients with acquired dyslexia (Coltheart, Patterson, & Marshall, 1987; Marshall & Newcombe, 1966, 1973) and later moving on to other areas, including spoken word comprehension (Ellis & Young, 1988). Although this approach is concerned with modelling healthy cognitive processes, much of the evidence that these models are built on comes from human case studies, or case series, consisting of patients with lesions and interesting associations, dissociations, or double dissociations, in terms of their cognitive profiles and behaviour. Marshall and Newcombe (1966) investigated two patients with reading impairments. Through analysis of their error patterns, they observed different profiles of impairments, which revealed two different subtypes of dyslexia: deep and surface level. Accordingly, they used the dual route model of reading to account for these patterns, and linked this to the neural level by speculating that damage to either processing route in the brain could lead to these distinct types of reading impairment. However, the cognitive neuropsychological approach does not make one-to-one correspondences between psychological processes and neural 'centres' in the brain: "Cognitive neuropsychological theories, in contrast, are nothing to do with the brain: they are the diagrammatic part of theories that specify how information is processed in language comprehension and production" (Howard & Hatfield, 1987). The assumption is that lesions can disrupt any of the modules, leading to particular patterns of deficits. Some deficits are more common than others due to the brain's functional architecture. However, because of individual variability in lesions to grey and white matter, no two people end up with identical patterns of impairment.

Later, Ellis and Young (1988) put forward a cognitive model of spoken word comprehension.

This included four key modules: auditory analysis (phonemes are extracted from the acoustic speech wave), auditory input lexicon (extraction of word forms), semantic system (extraction of word meaning), and speech output lexicon (spoken word forms). Through their investigation of patients with word repetition impairments, they speculated that three possible routes of processing must occur to account for the patterns of observed deficits, made possible by the connections between different modules which enables processing to bypass certain modules. In this way, damage to one module may allow an alternative part of the system to remain functionally intact. Functionally modularity is one of the defining principles of cognitive neuropsychological accounts of cognition. This framework still forms the basis of more recent cognitive neuropsychological models, as depicted in Figure 1-1. This model has been expanded to incorporate all domains of speech and language processing, and is based on a wealth of patient data. Today, this forms the basis of neuropsychological testing of speech and language impairments in patients, and a number of standardised assessments are available which tap into these processing routes, and identify particular levels of breakdown.

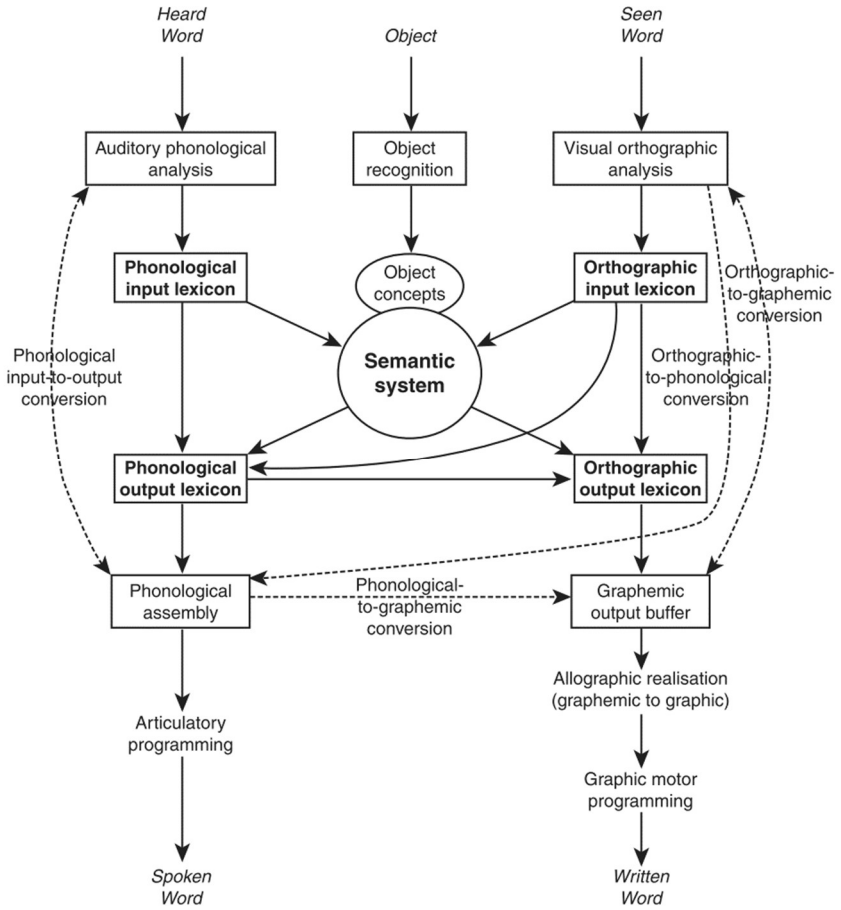


Figure 1-1 Cognitive neuropsychological model of single word processing, taken from (Whitworth, Webster, & Howard, 2014).

Figure 1-2 shows a subsection of this model for spoken word comprehension. A number of different patterns of impairments have been observed in patients with speech comprehension deficits:

- Impairments in auditory phonological analysis present as difficulty discriminating speech sounds, typically referred to as word-sound deafness (Best & Howard, 1994; Franklin, 1989). Impairment at this stage affects all later stages of processing, therefore patients are impaired in all repetition tasks (word and non-word) as access to the phonological input lexicon is blocked. PWA display difficulty on all speech input tasks, such as word-to-picture matching, synonym judgement, lexical decision, repetition and minimal pair discrimination (e.g. hair, pear). However, access to semantics from other domains may be intact, therefore reading could remain intact. Very rarely, this can occur as an isolated deficit

without other aphasic symptoms (pure word deafness) (Robson, Davies, et al., 2012).

- In word form deafness, individuals have difficulty recognising words as familiar, but do not have difficulty in analysing speech sounds, suggesting damage in access to, or within, the phonological input lexicon, where word forms are processed and/or stored (Franklin, 1989). Patients may have difficulty repeating real words, but non-word repetition remains intact due to the direct connection between auditory phonological analysis, and phonological assembly.
- In word meaning deafness, damage is in access between recognition of the word form (phonological input lexicon) and its meaning (semantic system). Individuals have good auditory lexical processing in other domains (e.g. they can read and/or write words) but experience impaired recognition of words in the auditory domain (Francis, Jane Riddoch, & Humphreys, 2001). This type of impairment is rare on its own, and is usually part of a constellation of other aphasia symptoms.
- Lastly, impairments at the level of the semantic system are typically attributed to difficulties with 'access' rather than 'storage'. This may vary depending on the type of task, or context. For example, there may be difficulty in resolving semantic competitors, or difficulty initiating links from sensory input (e.g. spoken or written word) (Thompson, Robson, Ralph, & Jefferies, 2015).

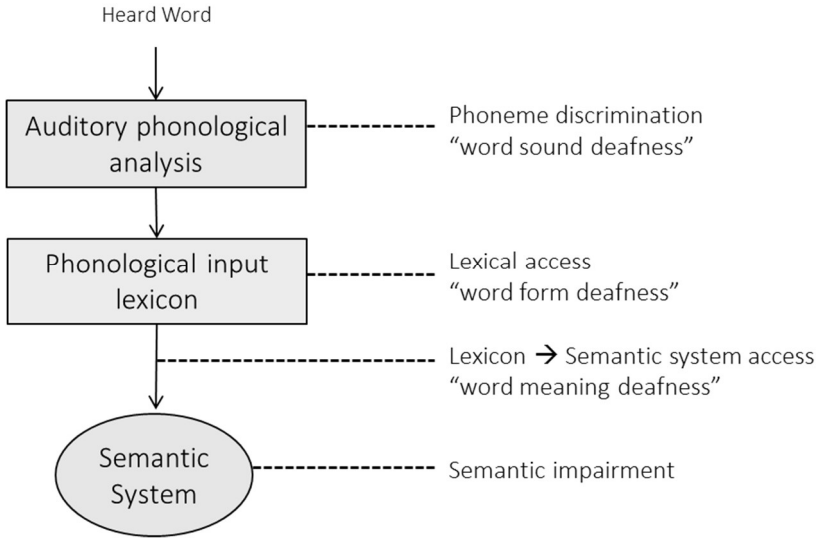


Figure 1-2. Cognitive neuropsychological model of spoken word processing (adapted from Whitworth et al. 2014).

1.2.2 Speech comprehension impairments and aphasia subtypes

Speech comprehension impairments are the defining feature of Wernicke's type aphasia (WA) (Goodglass & Kaplan, 1972). Patients have fluent spontaneous speech with a variety of grammatical constructions, normal articulation, rate, prosody and utterance length. However, speech is composed of paraphasias and/or neologisms, and 'empty' words which lack meaning (e.g. "thing"). Speech comprehension is moderately to severely impaired, and repetition is also affected. These classic features of WA can be seen in the case study by Ellis, Miller and Sin (1983) of patient RD. He showed poor speech comprehension, but good reading comprehension and spelling for words he couldn't say, indicating a particular impairment in processing spoken language. His spontaneous speech was characterised by many neologisms and some verbal paraphasias, indicating access to partial phonological information. In the same way, his spelling errors showed partial orthographic retrieval. The authors relate his level of impairment to problems retrieving the phonological specification of the word. In the current model, this may align with auditory phonological analysis and the phonological input lexicon.

Although classically associated with WA, speech comprehension impairments can also be seen in other aphasia syndromes, including Broca's aphasia, global aphasia, and conduction aphasia. Broca's type aphasia is characterised by non-fluent, agrammatic, and 'telegraphic' type speech (Goodglass & Kaplan, 1972). Spontaneous speech is characterised by omission of function words and grammatical affixes, simplification or omission of grammatical constructions, and predominance of nouns over verbs. Word retrieval problems are usually observed. Speech comprehension can be impaired but usually for more complex sentences, and repetition can be impaired, in line with speech output difficulties. Global aphasia is associated with extensive damage to the left (or language dominant) hemisphere. Individuals present with difficulties across all modalities, and can experience severe speech comprehension impairments. As previously described, individual patterns of lesion damage mean patients rarely present with the same precise profiles of impairments. As such, many PWA do not fit neatly into these subtypes, and can experience speech comprehension impairments as part of a constellation of aphasia symptoms.

1.2.3 Levels of breakdown in speech comprehension

In the model previously described, this first three levels of impairment relate to difficulty in discriminating the phonological aspects of speech. Within the literature, these have typically been referred to as auditory or phonological discrimination deficits. Accordingly, it has been well established that many patients with aphasia display impairments in both phonological discrimination and identification ability (Baker, Blumstein, & Goodglass, 1981; Blumstein, Baker, & Goodglass, 1977; Franklin, 1989; Maneta, Marshall, & Lindsay, 2001; G Miceli, 1980; Morris, Franklin, Ellis, Turner, & Bailey, 1996; Woolf, Panton, Rosen, Best, & Marshall, 2014). Given the modular and serial account of speech processing put forward, the logical assumption is that impairments at an early stage of processing will lead to downstream processing impairments, resulting in difficulty comprehending speech. This is in line with classic accounts, where "decoding deficits" were thought to be the cause of impaired comprehension (Luria & Hutton, 1977). However, the nature of the relation between phonological discrimination and auditory comprehension is unclear, and has been a matter of much debate. This is because patients have been reported who show dissociations in phonological processing and auditory comprehension (Baker et al., 1981; Blumstein et al., 1977; Gabriele Miceli, Silveri, Nocentini, & Caramazza, 1988), therefore the extent to which phonological discrimination impairments impact auditory comprehension has been questioned.

In a series of more recent studies of patients with classic WA, Robson and colleagues provide evidence which supports the presence of both non-verbal and phonological impairments in these patients, as well as a direct relation to speech comprehension ability (Robson, Davies, et al., 2012; Robson et al., 2013; Robson, Pilkington, Evans, DeLuca, & Keidel, 2017; Robson, Keidel, et al., 2012). In a case-series of 11 WA patients, the authors set out to extend the original work of Blumstein and colleagues (1977) by testing the relation between phonological discrimination and auditory comprehension (Robson, Keidel, et al., 2012). An adaptive consonant-vowel-consonant non-word discrimination task was administered, and thresholds were correlated with performance on a combined auditory comprehension score (which included words and sentences). In line with previous studies, patients showed severe impairments in the discrimination task

compared to healthy adults, and crucially, in contrast to Blumstein and colleagues (1977), thresholds correlated with impairment at both the spoken word and sentence level. This study improved on previous work by testing a much larger number of patients, and using a tightly controlled task which measured individual thresholds in phonological processing, and therefore provided a more sensitive measure of the degree of impairment. In a further group study with 10 patients, the authors used a number of auditory assessments sensitive to the spectral and temporal properties of non-verbal sounds (Robson et al., 2013). These included pure-tone discrimination, frequency modulation (FM), and dynamic modulation (DM). Group analyses showed impaired FM and DM detection, and thresholds for these detections at an individual level also correlated with auditory comprehension skills. These findings have led to the proposal that both spectro-temporal processing impairments and acoustic-phonological impairments are the cause of speech comprehension deficits in WA (Robson et al., 2013).

Others have suggested that semantic level impairments also contribute to comprehension failures (Baker et al., 1981; Miceli, 1980; Rogalsky, Pitz, Hillis, & Hickok, 2008). Robson and colleagues (2013) combine these findings by proposing that damage to both the phonological and semantic systems are likely to contribute to speech comprehension impairments in WA patients, due to lesion damage which extends across auditory and semantic regions, and the interaction between the two during speech comprehension. In other words, the intact system may compensate for the damaged system, but when both systems are impaired, large disruptions will take place. In this way, a small impairment in semantics may lead to severe speech comprehension impairments, in light of impaired phonological processing. More recent work provided support for this combined hypothesis (Robson et al., 2017).

Of note is that these findings are in contrast to Hickok & Poeppel's (2004) dual route model of speech processing previously described. In their model, speech processing up to the level of word recognition is bilaterally organised. WA patients with unilateral lesions are therefore expected to be able to carry out acoustic-phonological processing with the contralateral hemisphere, meaning deficits at this stage are not identified as the primary cause of speech comprehension impairments. Instead, impaired lexical-semantic

integration has been postulated as the level of breakdown, a process which is left-lateralised according to this model. Hickok & Poeppel (2007) suggest that bilateral acoustic-phonological processing may entail functional differences between the hemispheres, in terms of specialisations for different aspects of the auditory signal, a hypothesis also put forward by Vigneau et al., (2011) who suggest that bilateral RH activations may represent alternative aspects of processing, such as prosody.

These conflicting findings are not easily resolved, and are still open to investigation. However, they highlight that speech comprehension impairments are not a unitary symptom, but one which is likely to have multiple causes, and vary across patients. Conflicting views on the lateralisation of speech processing provides an interesting context for new research into the role of both hemispheres in speech comprehension impairments, as well as hemispheric contributions to recovery.

1.3 Neural basis of speech comprehension: evidence from lesion studies and healthy individuals

Impairments in understanding speech after stroke are typically caused by damage to the dominant temporoparietal cortex. Traditionally, lesions in the region of the pSTG have been associated with speech comprehension impairments in aphasic patients. However, recent neuroimaging evidence now demonstrates that word and sentence level speech comprehension is subserved by a widely distributed network of regions (Dronkers et al., 2004; Turken & Dronkers, 2011; Vigneau et al., 2006), therefore localising speech comprehension to one particular region is no longer considered plausible. More recent evidence suggests that the posterior STG is involved in phonological aspects of speech perception which facilitate articulation via sensory-motor integration. For example, Buchsbaum and colleagues (2011) investigated patients with conduction aphasia, who typically make phonological errors but have relatively intact comprehension, indicating an articulatory planning disturbance. Lesion analysis showed maximal overlap within the left posterior planum temporal. These findings are consistent with those who propose that the planum temporal functions within the dorsal speech processing pathway, as an auditory-motor transformation centre (Hickok & Poeppel, 2007; Warren, Wise, & Warren, 2005). Damage to this region would be predicted to result in speech output impairments, consistent with those observed by the conduction aphasic patients (Buchsbaum et al., 2011).

Lesion-symptom mapping studies investigate the relation between damage to particular regions in the brain, and resulting behavioural impairments, and have provided insight into the speech comprehension network beyond the pSTG. These approaches quantify the structural integrity of each voxel, and correlate this with a particular behavioural measure, such as spoken word comprehension. Two approaches are typically taken: voxel-based lesion symptom mapping (VLSM) and voxel-based morphometry (VBM) (Geva, Baron, Jones, Price, & Warburton, 2012). In VLSM, investigation is restricted to the lesioned area, and each voxel is assigned a binary value of ‘damaged’ or ‘not damaged’. In contrast, VBM employs a whole brain approach, and voxel values are continuous, reflecting the degree of structural integrity (Ashburner & Friston, 2000).

In the first study to report on this approach, Bates and colleagues (2003) investigated 101 chronic aphasic patients with speech comprehension impairments. Patients were divided into two groups depending on whether they had a lesion to each voxel. The authors found that damage to the posterior MTG was associated with speech comprehension impairments, as well as parts of the dorsolateral prefrontal cortex (DLPFC) and parietal association cortex. The dependent measure was a combined score which included word and sentence level comprehension, so the specific function of the MTG in auditory comprehension is unclear. However, in a separate VLSM study of 64 chronic aphasic patients, the MTG was as identified as being important for word level comprehension, whilst further frontal (mid-frontal cortex and IFG (BA46 and BA47) and temporal (pMTG, anterior STG, STS and AG) regions were associated with sentence level comprehension (Dronkers et al., 2004).

The MTG has been increasingly recognised as a central hub in word comprehension. Following on from the findings of Dronkers and colleagues (2004), Turken and Dronkers (2011) carried out structural and functionally connectivity of these identified regions in healthy participants. In structural analysis using diffusion tensor imaging (DTI), they identified the MTG region as a network hub, incorporating fibres from five major white matter pathways, which included the inferior occipito-frontal fasciculous, arcuate fasciculous (AF), inferior and middle longitudinal fasciculi, and the tapetum. In resting-state functional analyses, they observed connectivity with the left peri-sylvian association cortex, homologous regions in the RH, and association regions beyond classical language regions. As the MTG was previously associated with word level comprehension, the authors suggest that this may form a connectivity hub for integration between multiple sensory and motor domains which enable connections to be made between words and their meanings (Turken & Dronkers, 2011). A further VLSM study with 51 aphasic patients also suggests the MTG is important for multi-word integration during speech comprehension, and the authors postulate that breakdown in this function may be the cause of speech comprehension deficits in WA patients (Pillay, Binder, Humphries, Gross, & Book, 2017). This is in line with others who assert that speech comprehension

impairments in WA are due to post-phonemic breakdowns beyond the stage of word recognition (Hickok & Poeppel, 2000, 2007; Poeppel & Hickok, 2004).

The role of the MTG in word level comprehension is consistent with the evidence base for healthy speech processing. In a meta-analysis of functional imaging studies, DeWitt & Rauschecker (2012) found an 'anterior-directed processing gradient' in the LH temporal lobe, where mid-STG regions were most strongly active for phonetic aspects of speech processing, whilst anterior STG and STS were most strongly active for word-form and phrasal processing respectively. Anatomically, the STG/STS lies on the superior border of the MTG, therefore interactions between these regions is biologically plausible.

As described, Dronkers and colleagues (2004) found multiple regions in aphasic patients which were associated with sentence level speech comprehension impairments. These findings are in line with the complex nature of sentence processing, which is likely to involve the coordinated activity of multiple brain regions, with each region subserving particular aspects of processing, such as auditory short term memory (Leff et al., 2009). A follow up study investigating WM underlying these regions found the left MTG, anterior STG, STS, and IFG formed part of an interconnected network, with the MTG as a key connectivity hub (Turken & Dronkers, 2011). These regions were supported in a recent meta-analysis with healthy individuals, where activation likelihood estimation was utilised to combine activation foci from 45 different functional imaging studies investigating sentence processing (Walenski, Europa, Caplan, & Thompson, 2019). Results showed an extensive left-lateralised network of temporo-parietal-occipital clusters. Of note was a bilateral network in the temporal lobe, which included the pSTG, MTG, and temporal pole. However, as many of the included studies had low-level contrasts, it is not clear what function these RH activations are serving in relation to speech processing. Nevertheless, these findings demonstrate that sentence level processing entails a network of integrated brain regions. Damage to any part of this network, as demonstrated in lesion studies, is likely to disrupt this process and result in speech comprehension deficits.

More recently, Butler, Lambon Ralph and Woollams (2014) took a different approach by reducing assessment data, using principle component analysis, with behavioural data

from aphasic patients, to identify key dimensions and their neural substrates. They found that phonological skills were related to damage in the left perisylvian region, including the mid to posterior STG, STS, MTG and Heschl's gyrus, including WM corresponding to the dorsal pathway and the arcuate fasciculous. Semantics were related to the left anterior temporal stem (Butler et al., 2014). The advantage of this approach over others is that shared variance amongst individual assessments is factored out by generating components which provide a unique contribution to variance in the data. In this case, a clear separation was observed between phonological functions in mid-posterior temporal regions, and semantic functions in the anterior temporal lobe. These findings are consistent with research in patients with primary progressive aphasia, which has consistently found that severe word comprehension impairments are associated with atrophy in the ATL.

1.4 Recovery from Aphasia

A number of studies have demonstrated how patterns of lesion damage in the left hemisphere can lead to speech comprehension impairments (Bates et al., 2003; Dronkers et al., 2004), and also predict some of the variability in treatment outcomes in naming and reading (Aguilar et al., 2018; Fridriksson, Bonilha, Baker, Moser, & Rorden, 2010; Marcotte et al., 2012; Naeser et al., 1998). A related line of inquiry is identifying brain mechanisms which underlie recovery of function, whether due to spontaneous change, or therapy driven processes. By elucidating the neural correlates of recovery, it is hoped that treatments can be developed and refined which take advantage of these mechanisms, thereby promoting the efficacy of treatments for patients.

Recovery in aphasia is complicated by the neurobiological sequelae of stroke which can lead to further changes in brain structure and function over time. Wallerian degeneration occurs when a nerve fibre is damaged, leading to degeneration of the axon remote from the primary lesion site, away from the neuron's cell body. Diaschisis is a subsequent functional disruption of brain regions remote from the focal lesion site, due to deafferentation of neurons which were functionally connected to the lesion site. This leads to reduced metabolic activity in these regions. As a result of these processes, neural

tissue can be altered from regions remote from the lesion site, with associated influences on recovery.

1.4.1 Spontaneous recovery

A range of factors have been identified which may relate to aphasia recovery. These have included lesion and stroke related factors, including lesion location and size, type of stroke, aphasia severity and type of language impairment; patient related factors, including gender, pre-morbid cognitive deficits, age, education; and treatment related factors (Plowman, Hentz, & Ellis, 2012; Watila & Balarabe, 2015). In particular, lesion size and location and initial severity of aphasia have been consistently identified as key prognostic factors. After around six months patients enter the chronic stage of aphasia, which has classically been characterised as a plateau in language skills.

Hope and colleagues (2017) investigated longitudinal change in language skills, unrelated to any specific intervention, in 28 patients with left-hemisphere stroke, with a mean age at stroke of 52 years. Patients were followed up more than one year after their initial assessment session with behavioural assessments and structural MRI scans. There was significant variability in individual outcomes, with over half of the patients demonstrating improvements, but half showing declines in performance. Structural changes in the right hemisphere were found to systematically relate to, and predict, both directions of behavioural change, supporting a role of the right hemisphere in recovery. Specifically, increased grey matter in the right middle temporal gyrus (MTG) was associated with improved behavioural change, a region which was linked to word retrieval functions. No simple prognostic factors were found which could predict the direction of change (improvement or decline) or magnitude of change. These included demographic variables (age, gender, time since stroke), as well as structural variables (lesion volume, grey matter, and grey matter volume in regions which evidenced later change). This finding suggest that changes in language functions can continue within the chronic phase (I.E. language functions do not remain completely stable), but that the direction of changes is variable amongst individuals. However, some caution is required, as the sample in this

study is not likely to be representative of the stroke population, given the mean age of 52 years.

In sum, little is known about the long-term course of spontaneous recovery in the chronic phase after stroke. The study previously described suggests patients may show both gradual increases and declines in language skills in the years after stroke. In relation to the present study with chronic aphasic patients, sudden changes in performance are therefore likely to be due to treatment effects, rather than spontaneous changes in language skills. Currently, the PLORAS project (Predicting Language Outcome and Recovery After Stroke) is collecting behavioural and structural MRI data on a large number of stroke patients, to identify regions of the brain which contribute to long term language recovery (Price, Seghier, & Leff, 2010). This may be useful in the future for giving much need information to patients and their carers on the chances of recovery of particular skills, and also in identifying rehabilitation goals which are realistic for individual patients. Smaller scale treatment studies are still required which can shed light on the potential to respond to specific short-term therapies, in contrast to long term recovery.

1.4.2 Treatment related recovery and the role of brain structure

The accumulation of research in aphasia rehabilitation demonstrates that, on average, speech and language treatments are effective (Brady et al., 2016). However, no prognostic factors have been identified which consistently and accurately predict treatment success, despite studies which have carefully evaluated a wealth of treatment data over the last few decades (Bhogal et al., 2003; Robey, 1998). Identifying these factors is important, as aphasic patients show considerable variability in treatment outcomes, even across individuals with similar language profiles.

Previously, a small number of studies have investigated the contribution of pre-therapy brain structure to treatment outcomes in aphasic patients. A number of these studies have identified regions of damage in the LH which relate to treatment outcomes, in line with a link between lesion size and location and language recovery (Watila & Balarabe, 2015). In an early study, Naeser and colleagues (1998) investigated response to a

computerised visual communication system training program in 17 patients with post-stroke aphasia, using computed tomographic scans. They found that patients with bilateral lesions failed to learn from the program (N=4), and that for patients with unilateral LH lesions, specific patterns of lesions could predict two types of responders: those who could answer questions using the system (moderate responders, N=6); versus those who could use the system to also initiate communication (best responders, N=7). These findings demonstrate how lesion patterns may provide information about potential candidacy for types of treatment; however, initial behavioural testing was also able to accurately predict moderate and best responders, calling into question the clinical utility of evaluating brain scans in this case, as imaging did not provide a unique contribution to predicting outcomes.

Further studies have similarly identified regions of damage which relate to treatment outcomes; in these studies, behavioural and demographic factors were not reported to relate to outcomes. Marcotte and colleagues (2012) used a correlational analysis to investigate the relation between lesion damage and response to semantic feature analysis (SFA) naming therapy, in nine patients with chronic aphasia. Lesion size was not found to relate to improvements; however, damage to Broca's area (BA45) was strongly negatively associated with naming improvement, such that those who had more damage to this region tended to make the least improvements. Broca's area has classically been associated with language specific functions, including lexical selection and semantic encoding. In line with the authors interpretation, one possibility is that damage to this specific region impairs an individual's ability to improve on a task which engages these processes (I.E. semantic feature analysis). However, Broca's area is structurally heterogeneous, and in healthy individuals, subregions have been shown to be responsive to task difficulty during fMRI, regardless of type of stimuli (verbal or non-verbal) suggesting that the functions of this region are likely to be more complex (Fedorenko, Duncan, & Kanwisher, 2012). It is therefore not clear whether response to treatment in Marcotte and colleagues (2012) patients can be assigned to damage to specific language functions in Broca's area, considering the broader role of this region in cognitive processes.

In a similar study, Fridriksson, Bonilha, Baker, Moser, & Rorden (2010) investigated response to an anomia treatment in a group of 26 aphasic patients, but instead found that patients with damage to a specific temporo-occipital region in the LH were less likely to improve their naming ability following therapy, compared to those who had this region spared. These two studies assessed different patient groups (non-fluent and fluent respectively) and treatment tasks varied with respect to their use of semantic and phonological prompts, which may account for differences in the critical regions identified.

More recently, Aguilar and colleagues (2018) demonstrated that combining multiple variables may be one way to improve predictions about treatment response. The authors carried out a computerised treatment study (iReadMore), with 23 patients with reading impairments (central alexia) caused by left hemisphere stroke. Pre-therapy structural integrity was represented by proportion of lesion damage to 69 regions of interest (obtained using MRI scans), and these were combined into a model with further demographic and behavioural data. The winning model explained 94% of the variance in treatment outcomes, and included all three variable types, indicating lesion data could explain variation in treatment response more than behavioural and demographic data alone. However, as the authors acknowledge, the high degree of variance accounted for by the model indicates that overfit is likely to have occurred, in which the model is too close to the limited data points in the sample. In this case, the contribution of these variables to treatment outcomes in the population is likely to be overestimated. Accordingly, a second out of sample analysis predicted a much smaller percentage (23%) of the variance in treatment outcomes. With future larger samples, the predictive power of this model may increase.

In a novel study which looked at a region outside of the traditional language network, Meinzer and colleagues (2010) used MRI to investigate the integrity of the hippocampus in ten aphasic patients who underwent an intensive two-week language treatment. The hippocampus has previously been implicated in neural models of learning, for example, in paired-associate learning (Clark, Kim, & Maguire, 2018) and vocabulary learning (Breitenstein et al., 2005) in healthy adults, but its' role in aphasia treatment has not previously been explored. Using MRI methods, the authors found that patients with

lesions closer to the hippocampus, and more pronounced volume loss of the left HC (compared to their right HC), had poorer outcomes, providing preliminary evidence for integrity of learning and memory networks, as well as traditional language networks, in treatment outcomes in aphasic patients. This may relate to broader learning mechanisms not specific to language, as aphasic individuals also show variability in non-linguistic learning tasks (Vallila-Rohter & Kiran, 2013). Meinzer and colleagues (2010) patients did not have direct damage to their hippocampus, and so reduced volume is likely to be due to secondary atrophy. Studies which focus on damage in the lesion region may therefore exclude valuable information regarding integrity of further brain regions. One solution may be to utilise a whole brain approach which is sensitive to variations in volume caused by primary lesion damage, as well as secondary areas of volume variability, such as Wallerian degeneration. Voxel-based-morphometry is one such approach which has been shown to be sensitive to subtle regional variations in volume in aphasic patients which relate to particular linguistic and cognitive abilities (Butler et al., 2014; Leff et al., 2009b).

Further to structural imaging, and in line with these previous findings, a number of functional imaging studies have also demonstrated that integrity of functional networks prior to treatment may help to predict outcomes, in both hemispheres. For example, Richter, Miltner and Straube (2008) found that pre-treatment activation in the right IFG/insula during reading and word-stem completion tasks correlated with post-treatment success in their chronic aphasic patients, and also that success correlated with a decrease in activation. Similarly, Menke and colleagues (2009) found that short and long term training success was related to pre-treatment functional integrity of both memory (bilateral) and language (RH) related networks respectively. A further study also found that pre-treatment activity correlated with immediate treatment success, but in the LH, and also found that regions were different depending on the task (semantic/phonological) (Van Hees, McMahon, Angwin, de Zubicaray, & Copland, 2014)).

Taken together these findings demonstrate that both structural and functional imaging methods may provide much needed information about the likelihood to respond to particular treatments. Patterns of lesion damage may provide useful information for informing researchers, clinicians and patients about candidacy for particular therapies.

Currently, no study, to the authors knowledge, has reported structural or functional predictors for treatment success in therapies targeting speech comprehension, therefore little is known about the predictive value of brain structure and function to treatment outcomes in this domain.

1.4.3 Neuroplasticity and recovery of language functions

1.4.3.1 Evidence from healthy individuals

Advances in neuroimaging over the last thirty years have revealed how the adult brain is able to adapt in response to its environment, a phenomenon referred to as experience-dependent neuroplasticity (for reviews see Lövdén, Wenger, Mårtensson, Lindenberger, & Bäckman, 2013; Thomas & Baker, 2013; Zatorre, Fields, & Johansen-Berg, 2012). Two types of studies have typically been reported: cross sectional studies, which have compared 'expert' performers with novices, or groups of participants at different stages of skill acquisition; and longitudinal studies, which have tracked individual participants performance in response to particular training programs. During training programs, healthy individuals have often been tasked with carrying out specific and targeted practice of a particular skill. Findings may therefore inform us about the possible mechanisms which underlie language re-acquisition in aphasic individuals.

Some studies have reported greater regional brain volume in experts, suggesting that greater regional volume may support behavioural performance. For example, one study which investigated expert phoneticians found that amount of previous phonetic transcription training positively predicted the surface area of the pars opercularis, a region implicated in phonetic processing (Golestani, Price, & Scott, 2011). Other cross-sectional studies have investigated participants at different stages of expertise, and have similarly found a positive relation between brain volume and skill level. Mechelli and colleagues (2004) used VBM with individuals at different stages of second language proficiency, and found GM density in the inferior parietal cortex was greater in bilinguals than monolinguals, but also that second language proficiency positively correlated with density in this region.

Several longitudinal studies have demonstrated increases in GM density and WM integrity in response to various training tasks in healthy individuals. These have included learning to juggle (Draganski et al., 2004; Scholz, Klein, Behrens, & Johansen-Berg, 2009), learning golf (Bezzola, Merillat, Gaser, & Jancke, 2011), mirror reading (Ilg et al., 2008) and writing (Hamzei, Glauche, Schwarzwald, & May, 2012)). These studies have typically recruited young healthy participants, significantly younger than an average stroke population; however, one study demonstrated similar structural changes in elderly participants learning to juggle (Buchel, Gaser, Boyke, Driemeyer, & May, 2008).

A number of studies have also shown changes in the linguistic domain. Healthy adults have shown increases in local GM tissue density, or cortical thickness, in response to learning morse code (Schmidt-Wilcke, Rosengarth, Luerding, Bogdahn, & Greenlee, 2010), and learning a foreign language (Mårtensson et al., 2012), and in WM, changes in integrity have been observed in bilateral language networks in response to learning a new language (Schlegel et al., 2012). Of particular relevance for the present thesis is the finding that GM and WM can change rapidly in response to short-term language learning. Mårtensson and colleagues (2012) tracked performance, over three months, of interpreters taking part in an intensive language training program, which involved learning a large amount of vocabulary. Compared to controls, interpreters showed increased cortical thickness in left hemisphere language regions (IFG, MFG, STG) and increased volume in the right hippocampus. Furthermore, proficiency correlated with cortical thickness in the left STG and volume of the right hippocampus. These findings suggest that intensive word learning induces structural changes in regions implicated in phonological processing (left STG) and new word formation (hippocampus). In other words, treatment induced changes in regions specialised for language functions, as well as domain general regions important for cognitive functions which underpin learning.

More recently, Hervais-Adelman, Moser-Mercer, Murray, and Golestani (2017) investigated whether structural changes could be observed, using MRI, in interpreters undertaking 14 months of training. The authors found increases in cortical thickness in several regions which they attribute to linguistic and non-linguistic cognitive functions integral to being able to carry out simultaneous interpretation: lower-level phonetic

processing (left pSTG, anterior SMG and PT); higher-level speech formulation (right AG, right dorsal premotor cortex); and executive control and attention (right parietal lobule). In contrast, a control group showed decreases in volume of these regions over time, consistent with age-related cortical thinning. Of note is that the size of changes are small, particularly considering the extent of training which these individuals undertook over 14 months. For all regions, increases in cortical thickness were 4% or below; and for controls, decreases were in the region of -3%.

1.4.3.1 Methodological limitations in studies of training induced neuroplasticity

The small changes in brain volume previously described raise important considerations in studies of training-induced neuroplasticity. In a critical review of these studies, Thomas and Baker (2013) cite a number of methodological limitations which undermine the validity of some of these findings. These limitations include inadequate analysis of between groups studies, where a control group is compared to the training group (e.g. lack of statistical models which provide a clear group by time point interaction); absence of appropriate control training tasks; absence of control brain regions to evidence anatomical specificity; limited evidence for replicability across studies; and absence of correlations between behaviour and structural changes. Overall, Thomas and Baker (2013) make a number of suggestions for future studies to improve robustness of evidence.

In addition to study design considerations, Thomas and Baker (2013) also raise limitations relating to MRI. Human MRI is typically carried out with a $1mm^3$ voxel size; however, animal studies on which models of neuroplasticity are based typically observe structural changes at a much smaller microscopic level, therefore it is unclear whether these may be captured with MRI techniques in humans. For example, the authors cite two papers where cortical volume increased by .05mm and $.10mm^3$, values which are smaller than the sampling frequency used in these studies. Image processing steps can also introduce bias or false positives, such as the size of the smoothing kernel, and how voxels are aligned within and across participants (spatial normalisation) and how this is achieved across time

(longitudinal registration). Furthermore, the relation between probabilistic brain matter values (e.g. the probability that a voxel belongs to grey matter), which are the dependent measure used in VBM, and biological neuronal density, are not clear (Eriksson et al., 2009).

Thomas and Baker (2013) provide a comprehensive critique of the current evidence base. These considerations, and particularly the lack of replication of some findings, suggest that caution is warranted in interpretation of training induced neuronal plasticity. However, careful considerations of study design and processing techniques may increase robustness of evidence in future studies. In summary, the studies described previously provide some preliminary evidence for the presence of training induced structural neuroplasticity, including within the language domain; however, conclusions thus far should be taken with caution in consideration of important methodological limitations. Furthermore, these studies are based on healthy individuals, and it is not clear how these findings may extend to individuals with post-stroke aphasia, and associated neurobiological sequelae, such as Wallerian degeneration and accelerated age-related atrophy.

1.4.3.1 Evidence of neuroplasticity in aphasic patients

Few studies have reported longitudinal structural changes in aphasic patients in response to treatment. However, a wealth of functional imaging studies demonstrate changes in activation patterns, suggesting associated changes in structure are possible (Mohr, 2017). The majority of these studies have focused on speech output treatments. For example, improvements in naming have been associated with decreased activation in peri-lesional and homotopic regions overlapping with typical naming networks (Abel et al., 2015).

Two studies have reported structural reorganisation in aphasic patients utilising DTI. Diffusion tensor imaging is an MRI technique which uses the diffusion of water molecules to measure the microstructure of the brain (O'Donnell & Westin, 2011). This technique is based on the anisotropic properties of tissue, which results in variation in diffusion of water molecules with direction, and can therefore indicate the orientation of underlying tissue (I.E. white matter). Fractional anisotropy (FA) is a widely used anisotropic measure,

where reduced FA values represent greater diffusion, and therefore more well-define pathways. These values are typically interpreted as a measure of white matter integrity. This ratio measure describes the amount of tissue organisation, by identifying voxels which are likely to contain a single white matter tract (O'Donnell & Westin, 2011). Tractography techniques estimate the trajectories of white matter fibre tracts, and encompass a range of different methods, which as streamline tractography and probabilistic tractography (O'Donnell & Westin, 2011).

Schlaug, Marchina and Norton (2009) utilised DTI to investigate WM networks before and after MIT in six patients. The authors used a streamline tractography technique to re-construct white-matter tracts based on FA values from MRI. Following 75-80 daily session of MIT in six patients, there was a significant increase in the number of fibres in the right arcuate fasciculous, a key fibre tract linking fronto-temporal language regions. There were also significant improvements in speech outcome measures in all patients. The authors interpret these findings as "remodelling" of key regions due to long-term therapy, and suggest possible underlying neuronal mechanisms as increased myelination of existing axons, axon growth, or possible growth of more axon collaterals (parts of axons which provide connections with multiple synaptic targets). However, it is not possible to differentiate between these possibilities. However, specificity to treatment is unclear as there was only a non-significant 'trend' between volume change and behaviour.

In a further study, Wan, Zheng, Marchina, Norton and Schlaug (2014) found a reduction in FA values following MIT in 11 aphasic patients, in WM underlying RH language homologues, which included the IFG, pSTG, and posterior cingulum. Reduced FA values suggest greater white matter integrity as diffusivity increases along more well defined pathways. Unlike the previous study, FA values in the pars opercularis directly correlated with improvements in speech production, suggesting that changes in integrity are specific to treatment effects. These studies provide some preliminary evidence for modulation of WM architecture in aphasic patients, driven by therapy.

A paucity of studies have investigated reorganisation following speech comprehension treatment. In one study, PET was used to examine brain activation changes following a

short-term speech comprehension treatment in four aphasic patients with Wernicke's type aphasia, (Musso et al., 1999). The treatment involved 11 8-minute sessions performed during scan intervals. Training sessions involved five different tasks which focused on access to linguistic meaning (e.g. spoken command-to-picture matching), and increased in difficulty, resulting in gradual improvements in speech comprehension over sessions. Group results showed improvements correlated with activation in the right pSTG and left posterior precuneus, with additional areas which varied amongst individuals (in the left these were the dorsal frontal gyrus, IFG, cingular gyrus, lingual gyrus, and IPL; and in the right, MTG, IFG, MFG, SMG and postcentral sulcus). The authors interpret these findings in terms of homotopic compensation in Wernicke's area (pSTG), and memory related processes (precuneus and others). These early findings provided evidence suggesting reorganisation relating to treatment in both hemispheres, but have limitations due to the short-term nature of treatment, and small number of patients. Conversely, left and right IFG activity during semantic decision and verb generation tasks (rather than the temporal lobe), was associated with spontaneous improvements in sentence comprehension from 1 months to >1 year post stroke, in 13 aphasic patients (van Oers et al., 2010). However, the behavioural measure was the Token Test which entails significant working memory and executive processes, therefore activity may instead be related to increased processing demands, rather than speech comprehension per se. In a more recent treatment study, Woodhead and colleagues (2017) investigated activity changes, using magnetoencephalography (MEG), in 20 aphasic patients with speech comprehension impairments, following a high dose phonological training program. Improvements in speech comprehension (words and sentences) were associated with activity changes in the left STG, and for severe patients, in interhemispheric STG connections.

These studies provide evidence for neural reorganisation relating to speech comprehension. They suggest speech comprehension treatments may relate to changes in both hemispheres, and in line with typical language networks, point to both the temporal lobes for speech processing, and the IFG and further regions for processes which support speech comprehension. However, these conclusions are tentative given the paucity of research studies, and variability in methods and findings across individual

patients. No study has reported structural changes in response to speech comprehension treatment, therefore it is unclear whether structural changes may be observed in response to therapy. Based on the previous findings, if structural changes do occur, they may relate to both hemispheres, and in particular, regions around the STG.

1.4.4 Hemispheric contributions to language recovery

A pertinent theme in aphasia rehabilitation is the contribution of the left and right hemispheres. In a recent review of the evidence for aphasia recovery, Kiran and Thompson (2019) report inconsistent findings from fMRI research investigating treatment induced patterns of reorganisation. Whilst many patients (N=99) showed increased activation of LH regions, many patients (N=90) also show increased activity in the RH; however, the biggest proportion of patients (N=439) show increased activation in bilateral regions (Kiran & Thompson, 2019). Furthermore, the authors note that some studies report decreases in activation (e.g. Abel et al., 2015; Nardo et al., 2017)). These findings may not be surprising considering that LH speech comprehension regions are structurally and functionally interconnected with RH homologues (Turken & Dronkers, 2011), and that healthy individuals show bilateral activations across a range of language tasks (Vigneau et al., 2011). Furthermore, significant variability has been shown in lateralisation of WM language networks across individuals, suggesting hemispheric contributions to language processing are likely to be heterogeneous across the population (Catani et al., 2007). A number of studies have also suggested that hemispheric contributions may be dynamic in relation to short and longer-term outcomes (Saur et al., 2006).

The majority of previous treatment studies have focused primarily on speech output tasks, and therefore little is known about the contribution of the left and right hemispheres to speech comprehension recovery. The proposed bilateral nature of speech comprehension adds further to this debate. One possibility is that patients may be better able to engage RH networks to support speech comprehension. For example, according to Hickok and Poeppel's (2007) model, the RH should still be able to process speech input via redundancy of cues, and activate the mental lexicon. In functional imaging studies with healthy individuals, RH involvement was found to be lower in contrasts investigating

lexico-semantic processing (e.g. semantic association), but relatively higher in tasks engaging phonological processing and sentence/text processing (Vigneau et al., 2011). In terms of spatial distribution, activation peaks were gathered around the temporal lobe and two frontal lobe regions for lexico-semantic contrasts, the auditory cortex for phonological contrasts, and the superior temporal sulcus for sentence/text contrasts. If patients engage typical RH regions during speech comprehension treatment, then it is feasible that changes might be expected in these key regions. In light of mixed findings concerning the role of each hemisphere, and the relatively greater bilateral processing proposed for speech comprehension, it seems highly likely that both hemispheres may contribute to recovery in some respect. Studies investigating patterns of recovery may therefore hypothesise structural and functional changes in both hemispheres.

1.5 Behavioural Treatments for Speech Comprehension Impairments

1.5.1 Speech comprehension therapy tasks

Behavioural interventions for aphasia generally fall into five main schools: didactic, behavioural modification, stimulation, pragmatics, and cognitive neuropsychology (Howard & Hatfield, 1987). In clinical practice therapies draw on aspects from all schools, though cognitive neuropsychology is the most prominent approach in the UK. Assessments are used to identify specific areas of breakdown, and therapy is then planned accordingly (Whitworth et al., 2014). Unfortunately, few studies have examined speech comprehension treatments in aphasia, therefore only a few approaches to treatment have been reported (Basso, 2003; Whitworth et al., 2014).

The traditional approach for word comprehension impairments is spoken word-to-picture matching (Basso, 2003). In this approach, no distinction is made between levels of breakdown in auditory input (as previously described in the cognitive neuropsychological model of speech comprehension), and the semantic system. In other words, treatment is broad and may target a number of different loci of impairment. Phonological foils require a patient to distinguish between phonemes (e.g. tie/pie) and so target the auditory input

system, whereas semantic foils target the semantic system (Basso, 2003). These features may be manipulated to train different processes, or vary level of difficulty. Whitworth and colleagues (2014) describe a number of further treatment approaches which have been reported in treatment studies. Tasks have included phoneme-grapheme matching, phoneme recognition, syllable discrimination, same/different with consonant-vowel syllables, spoken word-to-picture matching, written word-to-picture matching, picture-word verification, and semantic categorisation (e.g. Maneta, Marshall, & Lindsay, 2001; Morris, Franklin, Ellis, Turner, & Bailey, 1996; Tessier, Agnes, Michelot, & Pascale, 2007). During tasks, individual parameters are manipulated to increase or decrease difficulty. For example, using a hierarchy of cues (Morris et al., 1996; Woolf et al., 2014), and focusing on a particular aspect of processing, such as using only semantic foils (Raymer, Kohen, & Saffell, 2006).

1.5.1.1 Findings from single case treatment studies

Compared to speech production, there is relatively little literature on therapy for auditory comprehension deficits in aphasia. A collection of single case studies have targeted therapy at the level of phonological discrimination. Morris and colleagues p.138 (1996) cite their study as "the first rigorous investigation of the efficacy of a treatment for auditory word comprehension problems that is based on widely recommended procedures". Their participant (JS), who had chronic aphasia and impairments in auditory analysis and semantics, improved on measures of phonological discrimination following training on six tasks (dose <12 hours). However, there was only a trend for improvements in comprehension (measured by lexical decision, synonym judgement, and sentence comprehension). Maneta and colleagues (2001) carried out a similar therapy program with their participant (PK), using three different tasks targeting phonological discrimination (dose was ~6 hours). PK showed no improvement in his phonological discrimination skills, in contrast to JS.

Tessier and colleagues (2007) gave computerised therapy tasks targeting phonological discrimination (phoneme discrimination and recognition) to a participant 10 months post stroke (dose was ~12 hours). They improved on measures of phonological discrimination,

and this generalised to auditory comprehension. However, this participant was unique in that they did not have any other aphasia symptoms, and had a lesion in their brain stem rather than cortex. Hessler and Stadie (2008, in Hessler, Jonkers, & Bastiaanse, 2010) also found generalised improvement for a range of phonological discrimination tasks in their participant, following a systematic program of phonological therapy.

Grayson, Hilton and Franklin (1997) used a within subject cross over design, comparing semantic therapy with a phonological-semantic therapy (I.E. multiple levels of processing). Their participant, LR, was six weeks post stroke and had phonological and semantic impairments. Semantic therapy consisted of spoken/written word-to-picture matching and categorisation tasks (dose was ~20 hours). Phonological-semantic therapy incorporated spoken word-to-picture matching with rhyming foils (dose was ~2.5 hours). LR improved word-to-picture matching with both therapies, but only the phonological-semantic therapy improved his minimal pair discrimination, suggesting specific therapy effects, and this generalised to untreated items. However, LR was one-month post stroke so some degree of spontaneous recovery is likely (e.g. his visual agnosia resolved in this period), and semantic therapy preceded auditory therapy which may have affected response to the latter.

Other studies have focused on therapy targeting semantic impairments, often via auditory comprehension. Francis and colleagues (2001) targeted therapy at access between the phonological input lexicon and semantic system, in their participant KW, who had word meaning deafness and was four years post stroke (dose: 3-4 times a week for 4 weeks). He had difficulty with tasks such as spoken word-to-picture matching, but had no phonological impairments. Therapy focused on the semantic aspects of words, such as reading definitions. KW showed significant improvements for treated items, but no generalisation to untreated items. Other studies have found positive effects for therapy targeted at the semantic system via auditory comprehension, with some also showing generalisation to untreated items (Bastiaanse, Nijboer, & Taconis, 1993; Behrmann & Lieberthal, 1989; Morris & Franklin, 2012; Munro & Siyambalapitiya, 2017; Raymer et al., 2006). More recently, a feasibility study found that two patients' spoken word comprehension improved following 40 hours of treatment, and improvements

generalised to untreated items (Knollman-Porter, Dietz, & Dahlem, 2018). This study reported significantly a higher dose than all previous case studies, providing preliminary evidence for the efficacy of high dose speech comprehension treatment. Of note is that this treatment differed from most previous case studies in that it used only one type of task (word-picture verification), therefore treatment was repetitive, and specific, which may have been conducive to promoting neural plasticity mechanisms (Kleim & Jones, 2008). Unfortunately, no other study has reported on a similar treatment.

1.5.1.2 Findings from group treatment studies

A number of studies have used group designs to explore phonological discrimination therapy. In a randomised controlled trial with 32 PWA, Prins, Schoonen and Vermeulen (1989) compared the effects of an experimental systematic auditory comprehension therapy, with "conventional stimulation therapy" (not specified further) and no therapy. No significant differences between any groups were found on a battery of language tests. The systematic therapy used 28 different tasks targeting auditory comprehension (from non-verbal sounds, up to morphosyntax), with each participant receiving approximately 40 hours. As the authors acknowledge, the broad range of tasks delivered over this therapy period meant no one component is likely to have received a sufficient dose. The details of the stimulation therapy were not elaborated on either, so it is hard to draw conclusions.

More recently, Woolf and colleagues (2014) evaluated both phonological and semantic therapy in a group of eight participants with varied aphasia profiles. The tasks involved phonological discrimination (of real words), with a hierarchy of cues, and the semantic condition also included verbal information on meaning. As a group there were no improvements in any outcome measure, and this was supported by individual performance across participants. One task which did show some improvement was picture-word verification. More recently, Knollman-Porter and colleagues (2018) utilised picture-word verification as a therapy task with two patients with aphasia, and found both patients made large and significant improvements for both treated and untreated items. Taken together with the positive findings from semantic based therapies previously

described, it appears that tasks which employ semantic processing may confer more beneficial therapeutic outcomes.

Two further group studies have reported positive findings for speech comprehension outcomes in aphasia. Archibald, Orange, and Jamieson (2009) investigated language outcomes from multiple domains in a pilot study with eight PWA, following a computerised therapy program (mean dose=22 hours). The program (AphasiaMate) contained a variety of tasks, which included 'auditory processing' modules (e.g. pointing to action pictures). They found significant group improvements on standardised measures of auditory comprehension, but in no other domains, demonstrating a specific effect of auditory comprehension therapy.

In a further study which used computerised training, Woodhead and colleagues (2017) delivered a variety of phonological training tasks (sublexical and word level) to 20 aphasic individuals using Earobics software, designed for adolescents and adults with developmental language disorder. They found a small but significant improvement on a standardised measure of speech comprehension, with an average dose of 72 hours. Of note is that this dose is considerably higher than the majority of previous studies (e.g. 6-12 hours, Maneta et al. (2001); Woolf et al. (2014)). However, the overall size of treatment effects were small, and so the clinical relevance of these for patients with aphasia is questionable.

1.5.1.3 Summary of current evidence for speech comprehension treatments

In summary, the evidence to date is sparse and remains inconclusive. The majority of studies have used either single-case or small group designs (<10 participants), and administered varying doses of therapy, making it difficult to generalise findings. Furthermore, amongst the patients who showed positive treatment outcomes, there are no clear behavioural prognostic factors which predicted treatment success, as they displayed varying aphasia profiles. For example, patient JS (Morris et al., 1996) and PK (Maneta et al., 2001) both presented with word-sound deafness as part of their global

aphasia, and underwent a similar structured program of therapy; however, JS made gains in phonological discrimination tasks, whereas PK failed to make any gains. One significant difference between these patients is lesion location; JS's was in bilateral white matter and left-hemisphere basal ganglia, whereas PK's lesion was confined to left-hemisphere temporo-parietal region. This highlights the possible influence of lesion site on treatment outcomes, which has not been directly investigated in the majority of aforementioned studies.

1.5.2 Impairment based treatment on a neuroscience basis

Recently, researchers in the field of aphasia rehabilitation have directed attention towards the neural basis of cognitive functions, and how treatments may capitalise on biological mechanisms which underlie learning (Kleim & Jones, 2008; Pulvermüller & Berthier, 2008; Varley, 2011). Damaged connectivity within and between neuronal assemblies which underlie language processing may be due to damage to the neurons themselves, or disconnection from each other (Pulvermüller & Berthier, 2008). In order to re-establish functional connections, experiences are needed which either strengthen existing synaptic connections, or establish new ones. Pulvermüller and Berthier (2008) describe two ways that learning in a neural system may occur. 'Coincidence learning' occurs when two neurons are frequently activated at the same time, allowing for connections to be strengthened. This is often referred to as Hebbian learning, where "nerve cells that fire together also wire together" (Pulvermüller & Berthier, 2008). 'Correlation learning' refers to the timing of activations, so that cells which fire independently of each other do not become associated. These two principles mean that the strength of the correlation in activations between two cells determines the strength of the connection. In relation to the learning experience during aphasia therapy, this translates to the simultaneous and repeated exposure of two objects, so that they become functionally connected with each other. For example, an auditory object (e.g. the word form), and a visual/semantic object. Based on these assumptions at the neural level, Pulvermüller and Berthier (2008) (p.566) put forward the following massed practice principle:

"It is advantageous to maximise quantity (number of therapy hours) and frequency (number of therapy hours per time) of language therapy"

This statement incorporates two important parameters of aphasia treatment which are still very much debated today. The first is the total amount of therapy a patient undertakes, typically referred to as dose (although this can also refer to other measures such as number of therapy trials etc.). The second is intensity of treatment, which is the frequency of treatment. These refer to two distinct principles; however, only recently have researchers attempted to disentangle the unique contributions of both to treatment (Dignam, Copland, et al., 2016; Dignam, Rodriguez, & Copland, 2016). Frequently, these measures have been conflated, or not reported, meaning the most beneficial regimens of particular therapies have not been identified.

1.5.2.1 Principles of neuroplasticity

In addition to dose and intensity, Kleim and Jones (2008) put forward 10 principles based on neuroscience research which promote experience-dependent neuroplasticity, specifically in the damaged brain. These include use and improvement of the function (versus, for example, learned non-use), training which is specific to the function being targeted, sufficient repetition, timing of plasticity in relation to training, salience, age, transference to other similar behaviours, and interference with acquisition of other behaviours. Perhaps the most pertinent for designing impairment-based treatments are dose, intensity, specificity and repetition.

According to Kleim and Jones (2008), activation of the neural pathway during a task is not sufficient to promote neuroplasticity. Reviewing the literature on learning in rats, they suggest that mass repetition is required to instigate neural change, noting that rats show changes in synaptic properties, but only after several days of training. In addition to instigating neural change, repetition is also required to make newly formed or strengthened connection resistant to decay, in the absence of training. This is similar to Pulvermüller and Berthier's (2008) proposition, that highly frequent training is needed to avoid uncorrelated activation weakening newly established or strengthened connections.

The specificity principle describes how specific kinds of neuroplasticity (e.g. localised to a specific brain region) depend on specific kinds of experience. In other words, training in one skill may not lead to enhancement of a similar skill. In terms of aphasia treatments, this implies that the target of treatment must be clearly defined and trained in a task which specifically taps that particular function. In addition, training must involve learning, rather than just use of the skill. In this respect, treatment tasks must consider how feedback will promote learning or acquisition of that particular skill. These principles of neuroplasticity are highly inter-dependent, and disentangling the contribution of one particular factor is a challenge. This is exemplified in the literature by studies which have often controlled for dose, but not intensity, or vice versa (Dignam et al., 2016). Furthermore, these factors must also interact with highly heterogeneous individual patient characteristics. The combination of factors which maximally support neuronal strengthening and reorganisation, for particular treatments, are therefore not yet understood, although dose and intensity have received particular attention in recent years.

1.5.2.1 Treatment dose

A number of reviews have now demonstrated the beneficial effects of high dose treatment, supporting the first component of the massed practice principle (Bhogal et al., 2003; Brady et al., 2016; Breitenstein et al., 2017; Cherney, Patterson, & Raymer, 2011; Raymer & Kohen, 2006; Robey, 1998). Bhogal and colleagues (2003) found that studies which delivered, on average, 98 hours of impairment-based intervention produced meaningful improvements in patients' language abilities. The most recent Cochrane review of Speech and Language therapy in post-stroke aphasia also concluded that there is a benefit of therapy when a dose of between 60-208 hours is delivered (Brady et al., 2016).

1.5.2.2 Intensity of treatment

In their massed practice principle, Pulvermüller and Berthier (2008) recommend highly frequent treatment, based on the underlying neural principle that highly correlated experiences are required to counteract uncorrelated experiences away from the therapy setting, which weaken the synaptic connections. In support of this claim, they refer to a parallel between groups treatment study which administered two regimens of treatments to chronic aphasic patients (Pulvermüller et al., 2001). Ten patients received an intensive treatment over 10 days (constraint induced aphasia therapy (CIAT)) whilst 7 received conventional treatment over 4 weeks. Overall, CIAT led to large and significant improvements in communicative skills, whilst conventional treatment failed to find the same effects. These findings support the use of highly frequent massed practice treatment; however, as the comparison task was a different type of therapy, it is difficult to draw conclusions about the unique contribution of intensity over other aspects of the treatment program.

More recent work has found less conclusive results. In a parallel group study with PWA, 6 hours of therapy per week was designated as a 'low intensity' condition, compared to a group who received 16 hours of therapy per week (Dignam *et al.*, 2015). The authors found no overall advantage for the high intensity condition. More recently, a large randomised controlled trial of 158 people with aphasia in the chronic phase of recovery demonstrated the benefits of high intensity therapy (approximately 31 hours per week) (Breitenstein et al., 2017). These inconclusive findings are compounded by ambiguity in the literature concerning the number of hours per week which constitute a low, medium and high intensity therapy regime (Cherney et al., 2011; Dignam et al., 2016). Some have regarded 5-10 hours per week as moderately intensive, and sufficient for therapeutic gains (Stahl et al., 2018). Conversely, in their study, Dignam and colleagues (2015) regarded 6 hours of therapy per week as a 'low intensity' condition. More systematic investigations into optimal treatment parameters are clearly warranted. Although there are uncertainties surrounding intensity, it is clear that high dose therapy is associated with improved language outcomes in PWA. Converging evidence from other cognitive domains also supports the notion that therapeutic interventions following stroke should be delivered at a high dose and intensity (Lohse, Lang, & Boyd, 2014; Schneider, Lannin, Ada, & Schmidt, 2016; Veekmans, Nopp, D'Haese, & Moeltner, 2004). Bhogal and colleagues

(2003), in their meta-analysis of studies, suggest that 1-2 hours per day over 2-4 weeks may be optimal.

1.5.2.1 Errorless and errorful learning approaches

Errorless learning is an approach which attempts to reduce self-reinforcement of errors (Fillingham, Hodgson, Sage, & Lambon Ralph, 2003). It is based on the view that production of an error for a particular stimulus will strengthen the association and make errors more likely on the next occasion. Fillingham and colleagues (2003) put forward two types of errorless learning conditions: error elimination, and error reduction. In the former, errors do not occur, and in the latter, errors occur but are reduced. In errorful learning, errors are not controlled. Previously, Mcclelland, Fiez and Mccandliss, (2002) found that subjects learning phoneme discriminations performed equally well in an error-reduction condition, and an errorful condition with feedback (versus no feedback). The authors suggest that Hebbian learning may be combined, or modulated, by reinforcement learning, to promote learning.

1.5.2.2 Practical considerations in delivery of impairment-based treatments

Despite evidence and consensus that high dose therapy is beneficial, a PWA is unlikely to receive this optimal dose within current service provisions, given that the average person with post-stroke aphasia receives less than 10 hours of Speech and Language therapy in the UK (Code & Heron, 2003). In line with the evidence, The Royal College of Physicians in the UK have recently updated their guidelines to recommend the provision of at least 45 minutes of therapy a day for as long as a person is benefitting from it (Bowen, James, & Young, 2016). However, it is unlikely that PWA will receive this dose, particularly as much of the 'treatment' time is absorbed by other important aspects such as information exchange, and initial assessment. In addition to difficulties with service provision, the nature of the therapy itself can also be a significant barrier to people with aphasia. Repetitive, mass practice exercises, which are a fundamental component of impairment-based therapies, easily become dull and mundane, and can act as a barrier to initial uptake

and continued engagement, and a limited pool of stimuli can contribute to this boredom and frustration (Kurland *et al.,* 2014). Indeed, higher drop-out rates are often observed for more intensive high-dose therapies (Brady *et al.*, 2016).

1.6 Use of Technology in Aphasia Interventions

The use of software programs and apps in language rehabilitation has grown considerably over the last few years, and there is preliminary evidence that this method of delivery is effective for treating aphasia (Lavoie et al., 2017; Roches & Kiran, 2017; Zheng et al., 2015), in treatments targeting a range of language and communication impairments, including word retrieval (Kurland et al., 2014), sentence production and comprehension (Thompson, Choy, Holland, & Cole, 2010), gesture (Marshall et al., 2013), reading (Woodhead et al., 2018), and apraxia of speech (Varley et al., 2016). However, few studies have investigated digital treatments specifically targeting spoken word comprehension (Raymer et al., 2006), and none have investigated high-dose computerised treatment.

One of the biggest benefits of utilising digital technology is the potential to increase autonomy with therapy tasks. A significant barrier to delivering therapy at a high dose is the limited contact time available to clinicians and patients. A clear advantage of self-administered digital therapy is the ability to move this into the patient's own time, allowing clinicians to focus on other equally important aspects of therapy (*e.g.* functional interventions and education). However, to date, this potential has not been fully realised given that many existing applications benefit from a therapist setting up and overseeing the therapy program (Mallet et al., 2019; Palmer et al., 2012). A related and unexplored avenue is developing automated programs which track users progress and adapt therapy tasks accordingly, increasing patient autonomy still further. For example, Sentactics®, a computer program based on treatment of underlying forms, provides an automated program which guides users through key therapeutic steps, and provides response-dependent feedback within each trial (Thompson et al., 2010). However, the therapy does not adapt to individual patient performance. One future avenue in digital aphasia treatment is therefore exploration of more complex computer algorithms which adapt to

meet individual requirements, both within patients over time, and across patients with different needs.

A further barrier to self-administered treatments is generating sufficient patient motivation to complete a high dose, in tasks which can often be dull and repetitive. Woodhead and colleagues (2017) used a self-administered computerised treatment for phonological processing skills in aphasic patients, which they navigated through independently at home. Although average usage was high suggesting a good level of tolerance (~70 hours), there was considerable variability in self-reported dose. The participants using Varley and colleagues (2016) apraxia of speech program also self-administered treatment at home. Consistent with Woodhead's participants, there was high variability in self-reported usage (0-52 hours), but a comparatively modest dose was achieved (M=15-18 hours). The authors report that many patients commented on the repetitive nature of the task, and use of the program dropped from the first to the second period of the study (M=18 and 15 hours respectively). Repetition is an important therapeutic principle, and is central to massed practice approaches which aim to promote language gains through neuroplasticity. However, if high self-administered doses cannot be tolerated by patients, then the true efficacy of these treatments may be difficult to establish, and their clinically utility, even if effective, is questionable if many patients cannot achieve a sufficient dose. Digital interventions therefore need to be mindful of incorporating strategies which can keep a user engaged for a significant period of time. To boost motivation and increase dose, Varley and colleagues (2016) suggest incorporating social and gaming elements into digital interventions.

A further benefit of utilising technology is the ability to increase treatment fidelity and thereby ensure validity of treatment research findings. Dignam and colleagues (2016) highlight the need for research which "systematically evaluate parameters of treatment intensity, such as dose, frequency and total intervention duration". An advantage of digital treatments is the ability monitor these parameters automatically, as opposed to self-reported or therapist-reported dose which may be subject to inaccuracies (e.g. time spent on instruction rather than treatment). Furthermore, content of therapy can be strictly controlled and stop 'therapist drift', whereby therapist-administered protocols

can be unintentionally modified over time in response to individual patients. Digital therapies therefore offer a reliable method to record actual time spent on therapy, and keep the program consistent within and across users, enabling explicit and easy reporting of treatment fidelity.

1.6.1 Gamification of aphasia treatment

1.6.1.1 Principles of gamification

A potential solution to the monotony of mass practice tasks is digital gamification. Gamification involves adding gaming elements to an existing task: "The aim of gamification is to use game like features (competition, narrative, leader boards, graphics, and other game design elements) to transform an otherwise mundane task into something engaging and even fun" (Lumsden, Edwards, Lawrence, Coyle, & Munafò, 2016). In this way, the repetitive elements in an impairment-based task may become better tolerated by patients, by providing an overarching experience that engages and motivates users, to keep them practicing for longer periods.

1.6.1.2 Gamification of previous assessments and treatments

Within the cognitive domain, gamification has been applied successfully to a number of cognitive tests and training applications (for a review, see Lumsden et al., 2016). For example, a trial which included older adult participants (60-85 years), by Anguera and colleagues (2013), successfully implemented a cognitive control training task within a racing videogame ('Neuroracer'), demonstrating that gamification can be both suitable and feasible within the field of cognitive rehabilitation. Woodhead and colleagues (2018) recently developed and tested a digital training application for reading ('iReadMore') in PWA with central alexia. Several gaming elements were incorporated into the therapy task itself, and included reward points, goals, sound effects, and animations. iReadMore was well tolerated by patients who achieved a large average dose of 70 hours, as well as significantly improving their single word reading accuracy. To the authors' knowledge,

gamification has been otherwise unexplored in evidence-based speech and language therapy applications for persons with aphasia.

1.6.1.1 Considerations for gamification of therapy in a stroke population

A research study carried out by the British Broadcasting Coorporation (BBC) surveyed 3442 individuals in the UK, aged between 6-65 years, to investigate the size of the gaming market in the UK, and profile types of 'gamers' (Pratchett, 2005). Of the individuals surveyed, 59% were identified as 'gamers': "someone who had played a game on a mobile, handheld, console, PC, Internet or interactive TV at least once in the last 6 months" (Pratchett, 2005). However, for the subgroup of respondents aged 51-65 years, only 18% were identified as gamers. Given the average PWA is over 70 years old, many individuals with aphasia are likely to be inexperienced with gaming technology, highlighting the need for careful consideration of gaming requirements for this demographic.

A small number of studies have reported on game design for an older population (Ijsselsteijn, Nap, de Kort, & Poels, 2007; Marston, 2013; Nap, Kort, & IJsselsteijn, 2009). Older users have typically received less exposure to gaming technology, and may therefore have less developed 'mental representations' for digital technology and gaming (for example, having an implicit understanding of how gaming 'levels' function). This has implications for the usability of digital applications; for example, older adults have been shown to make more errors when using technology, and tend to make use of different strategies, such as being reflective rather opting for a trial and error-based approach (Nap et al., 2009). Less experience may also be related to confidence in using technology. Marquie and colleagues (2002) investigated feelings towards computers with older and younger adults, and found older adults had poorer computer related self-efficacy. One way of improving self-efficacy could be to build in rewards for 'mastering' the application (Marquié, Jourdan-Boddaert, & Huet, 2002), and so reducing the potential impact of this barrier on engagement.

A final consideration are normal age-related changes in sensory-perceptual, motor and cognitive processes. For example, Nap and colleagues (2009) report on two individuals (> 60 years) who were unable to complete levels in their respective games due to imposed speed and time constraints. In addition to these typical changes, patients with aphasia are likely to have a range of other cognitive and motor impairments which may affect their use of technology (Szabo & Dittelman, 2014). To appeal to a broad range of individuals with aphasia, and promote independent use, treatment apps are needed which provide an intuitive interface and design, without heavy cognitive or physical burdens. As impairment-based treatments are beneficial when delivered at a high dose, treatment apps also need to ensure that the user experience supports this level of uptake.

1.6.1.2 Co-design in development of digital technologies with aphasic patients

These considerations highlight the need to involve patients directly in development stages of treatment apps. However, despite the increase in digital applications for aphasia, few have reported on this kind of participatory approach. One study which did specifically investigate the process of co-design with individuals with aphasia, noted a number of challenges, such as differing level of language impairments (Wilson et al., 2015). However, their overall experience suggests that a collaborative co-design process is feasible for this population, if well thought out activities are developed which enable participation in light of language impairments.

2 Overview of this thesis

2.1 Chapter 1

This section relates to the development of the Listen-In therapy app, in collaboration with persons with aphasia and software developers. This iterative cycle of development focused on integrating the therapeutic component within an overarching game. A schematic of this development phase can be seen in Figure 2-1, showing the anticipated iterative cycle of development, in collaboration with software developers, and persons with aphasia. Alongside my colleagues, I developed the therapeutic component in parallel to this process, which was based on a traditional spoken-word to picture matching paradigm (a detailed description of this is provided in Methods). I adapted this traditional paradigm by incorporating a number of parameters to allow independent practice and automated and adaptive progression through increasingly difficult challenges based on individual performance. This aspect of the project also involved development of a large range of therapeutic stimuli.

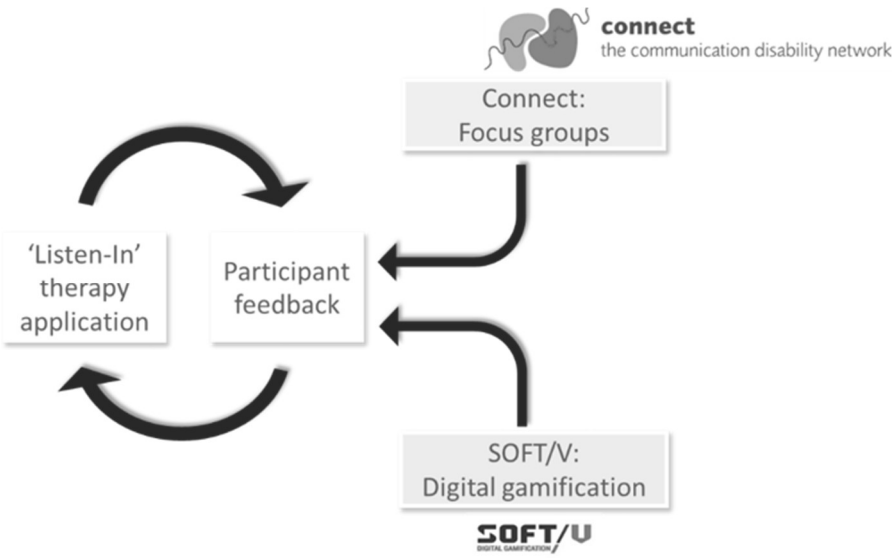


Figure 2-1 Development phase of Listen-In, showing an iterative cycle of development in collaboration with software developers, and persons with aphasia.

2.2 Chapters 2, 3 and 4

These sections relate to the randomised repeated measures cross-over trial of Listen-In (Figure 2-2). Thirty-five aphasic participants completed the trial, and were tasked with completing 100 hours of therapy over 12 weeks (therapy block), followed or preceded by 12 weeks of no Listen-In therapy (standard care block). Behavioural testing took place at five time points 12 weeks apart: baseline (T1), three intermediary time points before and after each cross-over block (T2, T3, T4) and a 12-week follow-up (T5). Structural magnetic resonance imaging (MRI) scans were performed before and after each cross over block (T2, T3, T4).

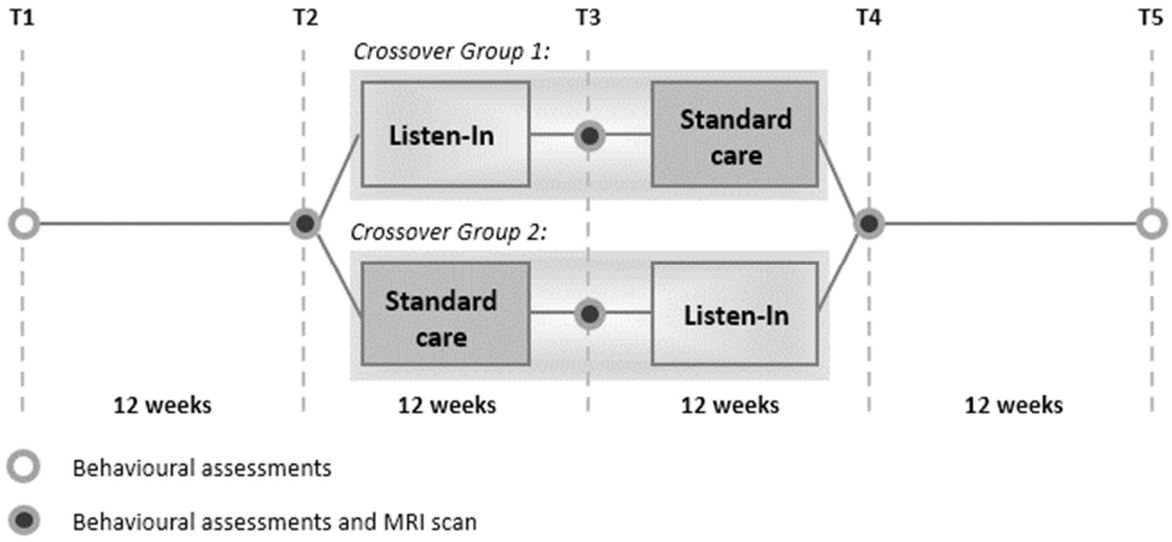


Figure 2-2 A flow chart demonstrating the randomised cross-over repeated measures trial design. Behavioural assessments were conducted at all time points (T1-T5) and structural MRI scans were obtained for a subgroup of participants at T2, T3 and T4. Listen-In=12 weeks of self-administered home-based Listen-In treatment. Standard Care=usual activities which the patient typically undertakes. T=time point.

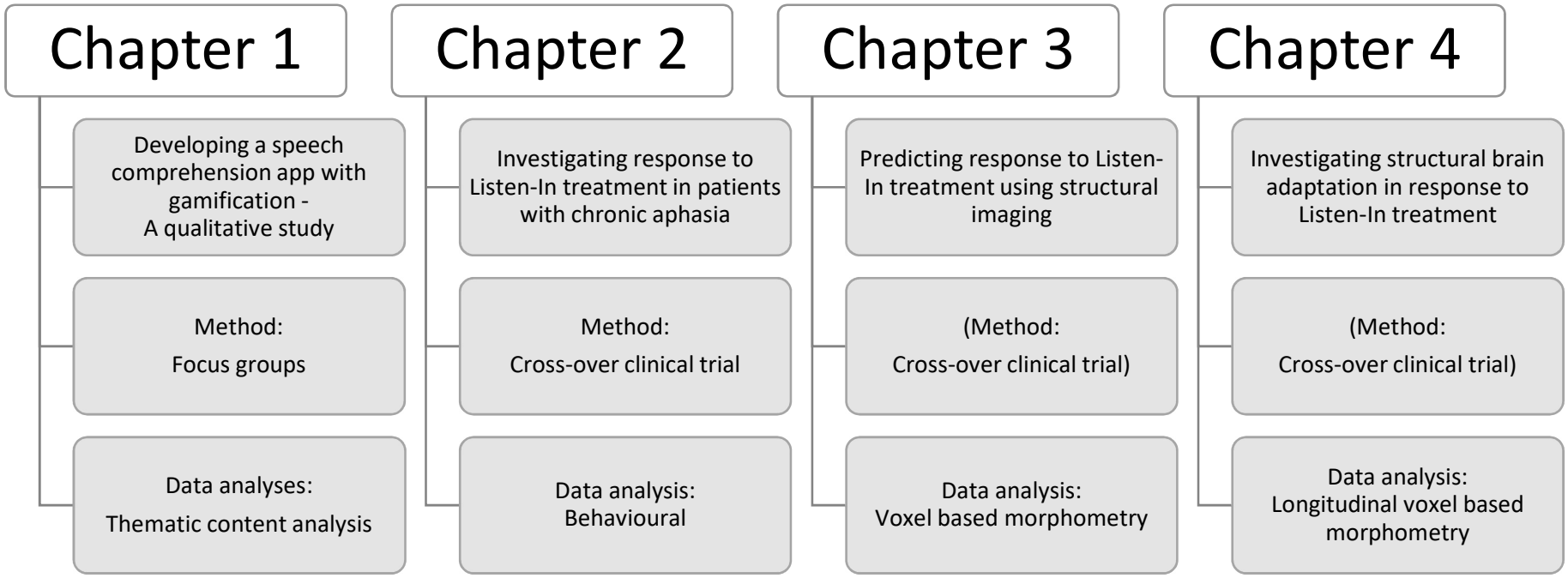


Figure 2-3 Details for chapters one to four.

3 Main Aims and Hypotheses

This results in this thesis are presented in four chapters (Figure 2-3). The aims and hypotheses of these chapters are outlined below.

3.1 Chapter 1: Developing a speech comprehension app with gamification - A qualitative study

Aim

To develop a speech comprehension therapy application ('Listen-In), with gamification, using co-design methods with persons with aphasia.

Rationale

Technology offers one way for patients to self-administer a high dose, from the comfort of their own home. However, therapies which entail mass repetition can become dull and boring, and be a significant barrier to achieving a high dose. Gamification offers a potential solution to this barrier, but this approach has not been previously applied to digital aphasia therapies. It is therefore unclear what factors may contribute to a gamified therapy app which is useable, and enjoyable, for this population. The present chapter reports on a qualitative study which aimed to develop a speech comprehension therapy app, with gamification ('Listen-In'), for use by aphasic patients in a follow-on clinical trial (Chapters 2-4).

Objectives

1. To carry out five focus groups with individuals with aphasia, to obtain feedback on a series of Listen-In prototypes
2. To collect qualitative data from five focus groups, and analyse this data to identify key issues surrounding usability and enjoyability

3. To develop design recommendations for each key issue, and implement these into subsequent iterations

Hypotheses

This study employed qualitative methodology, and as such, proceeded inductively, with the app being developed and refined during the process of data analysis. I expected to draw a number of themes from the qualitative data which would relate to usability of the app, and enjoyment of the app. I expected to translate these into tangible design changes, to produce a final product which could be used independently by persons with aphasia, but which was also sufficiently motivating and enjoyable to support high dose treatment. Given the exploratory nature of this study, I did not anticipate any particular usability or enjoyability issues.

3.2 Chapter 2: Investigating response to Listen-In treatment in patients with chronic aphasia

The literature previously reviewed shows a paucity of studies have investigated speech comprehension therapies, and these have shown inconclusive findings. Studies are highly heterogeneous, and have employed different treatment methods, small overall numbers of patients, and given small overall dose. In light of more recent findings demonstrating the benefits of high dose speech and language therapy, a larger group study is needed to ascertain whether, if given in a sufficient dose, traditional spoken word-to-picture matching therapy is effective at improving speech comprehension skills in persons with aphasia.

Aim 1

My main aim is to investigate whether Listen-In can improve comprehension of spoken words in patients with chronic aphasia. Specifically, I will investigate whether a large total dose of self-administered computer-based therapy can improve comprehension of treated and untreated spoken words.

Aim 2

My second aim is to investigate individual treatment effects, to identify patients at an individual level who respond to treatment.

Aim 3

My third aim is to explore whether baseline factors relate to treatment outcomes. For this aim, I will investigate whether a combination of factors can explain treatment outcomes.

Hypotheses

(1) At the group level, patients will significantly improve in their comprehension of spoken words trained during Listen-in treatment.

Not all speech comprehension studies have demonstrated treatment effects. However, I expected that given the high target dose and specificity of the task, that the majority of

participants would show improvements for treated items. Listen-In incorporated semantic and phonological components. According to the cognitive neuropsychological model of spoken word comprehension (Whitworth et al., 2014), this therapy may target all levels of auditory input processing, including: auditory phonological analysis, phonological input lexicon, semantic system, and access between these modules. I expected that repeated pairings between the word form and its meaning would strengthen the mapping between the phonological form and its meaning, resulting in improvements on treated items.

(2) At the individual level, some patients will significantly improve their comprehension of spoken words for both trained and untrained items following Listen-In treatment.

There is limited evidence available for generalisation in speech comprehension therapies. Listen-In incorporated two main phonological training components: exposure to the auditory word form (paired with its meaning), and phonological discrimination in the form of phonological foils. Theoretically, if therapy improves the early stages of speech perception up to word recognition (auditory input analysis) then improvements may be seen across all forms of speech (treated, untreated). However, given the broad inclusion criteria it is likely that many patients have impairments beyond phonological processing which impacts speech comprehension (e.g. lexical and conceptual semantics), therefore these improvements would only be expected for some patients.

(3) At the individual level, patients who significantly improve their comprehension of spoken words for both trained and untrained items, will show a significant improvement in their phonological discrimination ability following treatment.

Patients with speech comprehension impairments have been shown to have deficits in phonological processing, such as phoneme discrimination (Robson, Keidel, et al., 2012). As described in Hypothesis 2, it is possible that Listen-In training will improve some patients phonological processing skills, leading to improvements in untreated items. In this case, these patients may show concurrent improvements in phoneme discrimination.

(4) Patients will not show changes in non-verbal auditory processing skills.

Many patients with speech comprehension deficits also show concurrent deficits in non-verbal auditory processing (Saygin, Dick, Wilson, Dronkers, & Bates, 2003), therefore patients were expected to show impairments in the environmental sound discrimination task (ENVASA). However, it was not expected that Listen-In would improve general auditory processing skills, as this was not a target of the therapy task. Therefore this measure was not expected to show significant changes.

(5) Baseline demographic and behavioural factors will explain a proportion of the variance in treatment outcomes.

Previous case and group studies have shown variability amongst patients in magnitude of treatment effects, as well as variability in whether patients responded to treatment or not. In line with these studies, and aphasia treatment research in general, I expected patients to show variability in speech comprehension outcomes. Previously, in one study, a combination of baseline factors were shown to explain some of the variance in treatment outcomes in patients with central alexia (Aguilar et al., 2018). Given the range of baseline measures collected, I expected that a combination of these factors may account for some of the variability in response to treatment in the present study. However, as previous studies have failed to find consistent predictors, I did not have any specific hypotheses for what these baseline factors may be.

3.3 Chapter 3: Predicting response to Listen-In treatment using structural imaging

The literature previously reviewed demonstrates that structural brain imaging can be used to help predict therapy driven outcomes. However, these studies are rare, and none have investigated the role of baseline brain structure in relation to speech comprehension treatment. This chapter explores the relation between baseline brain structure and changes in spoken word comprehension following treatment (Listen-In).

Aim

To investigate whether structural integrity of pre-treatment grey matter (GM) and white matter (WM) predicts response to Listen-In treatment.

Hypotheses

(1) Structural integrity of baseline GM and/or WM in the left hemisphere (LH) will correlate with change in spoken word performance from pre to post treatment.

(2) If significant correlations are found, these may be observed in the left hemisphere (LH) language network, in peri-lesional regions, where variability in integrity will be present amongst the patient group.

A handful of previous studies have identified lesion patterns in the left hemisphere which predict, in part, variability in treatment outcomes (Aguilar et al., 2018; Fridriksson, 2010). My hypothesis was based on these findings. I expected to find peri-lesional regions in the left-hemisphere, particularly around the superior and middle temporal gyri, regions implicated in phonological and lexical processing. I expected to find a positive relation linking greater volume with greater treatment outcomes. My rationale for this hypothesis is that patients with greater damage to key speech comprehension regions will have reduced neural processing ability to support behavioural changes in speech comprehension.

3.4 Chapter 4: Investigating structural brain adaptation in response to Listen-In treatment

Findings in the expert performance literature with healthy individuals demonstrate that structural MRI methods can be sensitive to subtle differences in regional brain structure relating to particular skills. A number of longitudinal studies have also demonstrated changes in brain tissue in response to skill acquisition. However, there is little evidence to date which has investigated this in persons with aphasia, in response to particular interventions. As such, this chapter focuses on investigating whether structural neuroplasticity can be observed in persons with aphasia in response to Listen-In.

Aim

To investigate whether regional changes in brain tissue are observed in relation to Listen-In treatment, in patients with chronic aphasia.

Hypotheses

Listen-In therapy shares many similarities with previous training studies in healthy adults, which typically involve massed practice repetitious tasks which focus on one particular skill. However, the current study investigated this question in individuals with lesioned brains and disrupted brain networks relating to the skill being trained, therefore it was not clear whether changes could be observed at a group level, given possible heterogeneity in functional and structural reorganisation, both prior to therapy, and as a result of therapy. Nevertheless, the following hypotheses were identified:

(1) Improvements in speech comprehension skills will correlate with tissue changes in the temporal lobe in the LH, a key region of the speech comprehension network.

I expect that the high dose and repetitious nature of Listen-In treatment over 12-weeks will stimulate the residual speech comprehension network in the LH and lead to changes in neural tissue which relate to treatment outcomes. Previously, studies have shown that these peri-lesional and peri-sylvian regions are involved in language recovery, therefore

this hypothesis is in line with this literature. Finding tissue changes in this region would support evidence which suggest the LH plays a facilitatory role in treatment related recovery (e.g. Meinzer et al., 2008).

(2) Improvements in speech comprehension skills will correlate with tissue changes in the temporal lobe in the RH, in bilateral speech processing regions.

There is significant debate in the literature on the role of the left and right hemispheres in aphasia recovery, including therapy driven recovery. There is also a paucity of speech comprehension studies which have investigated treatment induced neural reorganisation. Two previous studies have suggested that treatment may involve changing activations in both hemispheres (Musso et al., 1999; Woodhead et al., 2017). This is in line with the proposed bilateral organisation of speech processing (Hickok & Poeppel, 2007). For these reasons, my second hypothesis was that if changes were observed in relation to treatment, they may also be observed in the RH. Finding tissue changes in the RH, which relate to improvements in speech comprehension, would support evidence which suggests the RH can play a facilitatory role in treatment related recovery (Cocquyt, De Ley, Santens, Van Borsel, & De Letter, 2017).

(3) Total exposure to auditory stimuli over the Listen-In treatment block will correlate with tissue changes in bilateral temporal lobes, in the speech processing network.

Previous findings in experience-driven neuroplasticity have often reported changes associated with amount of training, rather than behavioural change. A number of studies have shown that the human auditory cortex shows task-dependent neuroplasticity, such as during pure tone discrimination (Ohl & Scheich, 2005). Given the high number of auditory stimuli which patients were exposed to in treatment, I expect that changes in tissue may therefore be observed in auditory speech processing regions in bilateral temporal lobes. To investigate this hypothesis, I will use dose (hours) as the independent variable.

4 Methods

4.1 Chapter 1: Development of Listen-In

4.1.1 Qualitative research approach

Focus groups were chosen to investigate the aims of Chapter 1, as they were well suited to explore the Listen-In prototype in two ways: (1) By assessing usability across multiple participants quickly and efficiently, through direct observation of participant interactions with the tablet and app, and; (2) By enabling small group discussion around the app on pre-selected topics, as well as allowing room for spontaneous and unanticipated responses to emerge via a shared social setting.

The groups in this study were conducted in line with participatory style focus groups, the characteristics of which are described in Table 4-1, based on recommendations set out by Krueger & Casey (2015). Wilson and colleagues (2015) have also set out recommendations for the co-design process, specifically with persons with aphasia. Table 4-1 summarises these recommendations, and describes how these methods were applied to the planning and design of focus groups in the present study. Following the first group, activities and approaches were adapted in a flexible manner to continually meet the needs of the researchers and persons with aphasia, described further in Chapter 1. See Appendix 1 for an example of the focus group format.

Recommendations	Methods used in present study
Krueger and Casey (2000) – Key recommendations for focus groups	
Number of participants 5-8 participants, more if the purpose of the study is to pilot-test materials. Participation in multiple groups recommended if investigating change in perceptions over time.	12 participants recruited 5 to 8 participants in each group All participants invited back for all groups, to assess changing reactions to Listen-In iterations.
Recruitment Many strategies recommended depending on target group.	Organisational recruitment chosen with charity Connect. Advantages include: - More likely to participate due to existing relationship and trust with Connect - Ready-made pool of potential participants - Participants familiar with, and able to travel to, location - Trust already developed amongst members
Moderators Team of moderates recommended to perform specific tasks.	One facilitator designated for each group or subgroup. A team of moderators experienced with persons with aphasia took part to support communication.
Question route A pre-planned question route guides discussion through a sequence of specific questions. Around 12 questions recommended for two-hour group.	Question route developed for each group. Questions kept simple and supplemented with visual aids to support comprehension.
Capturing discussion To capture group discussion use: field notes, rating sheets, audio recordings, on-line transcripts, video recordings. Video recording not recommended due to potential intimidation effects, and single view of group.	Video recordings, rating sheets, and field notes were used across all groups. Although not recommended, video recordings were used as the primary method to enable non-verbal communication to be recorded in full, for later analysis, considered important due to the population. Two cameras were used to capture different aspects of the group.
Report Findings (typically organised around key themes or ideas) Researcher recommendations (what action should be taken from the findings)	Report organised according to key findings and recommendations. Recommendations were concrete design changes for Listen-In, based on key findings.

Wilson and colleagues (2015) – Co-design recommendations for persons with aphasia	
Prototypes Use of incomplete versions of a product	Prototypes of the app formed the basis of all focus groups, and were the most prominent method employed. Participants were asked to use the prototypes at home as if they were carrying out therapy, providing direct experience and engagement without the need for verbal exchange. As the prototypes were not fully developed, this enabled participants to take on the role of co-designers.
Images Presenting concepts visually rather than verbally	Screenshots and other visual represents were developed for all anticipated topics in the questioning route (e.g. Figure 4-2).
Total communication strategies Incorporating a range of strategies from all domains. E.g. text, gesture, pictures	Total communication strategies were carefully considered before each focus group. A key part of this was developing visual tools to provide a common reference point. This enabled shared meaning, and facilitated total communication.
Visual usability measures	Topics and questions were supported by visual options. For example, if participants were asked which option they preferred, visual screenshots of the options were developed, alongside key text and verbal support, to facilitating understanding of the concept (e.g. Figure 4-1).
Demonstrations Demonstrating as a means of explanation.	A key part of these focus groups was assessing usability. For this reason, demonstrations were provided only if a participant was unable to use the app after a period of independent play.

Table 4-1 Key focus group recommendations, and methods used in the present study design. Taken from Krueger & Casey (2000), and Wilson and colleagues (2015).

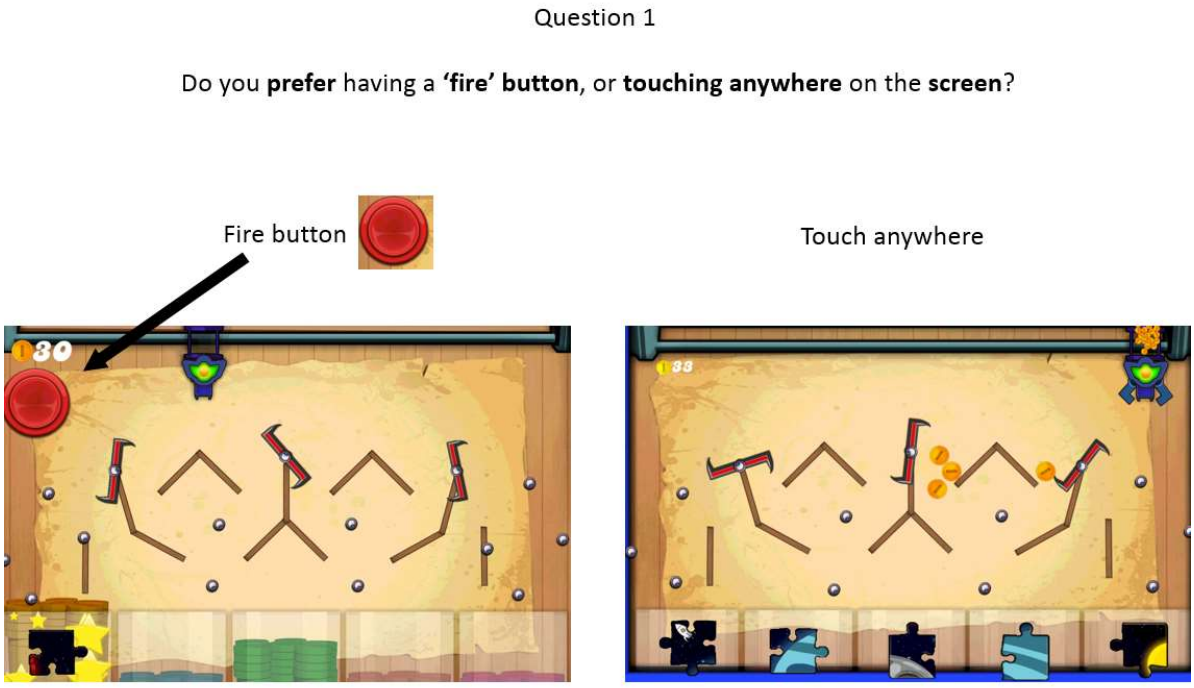


Figure 4-1 Example of visual usability measure

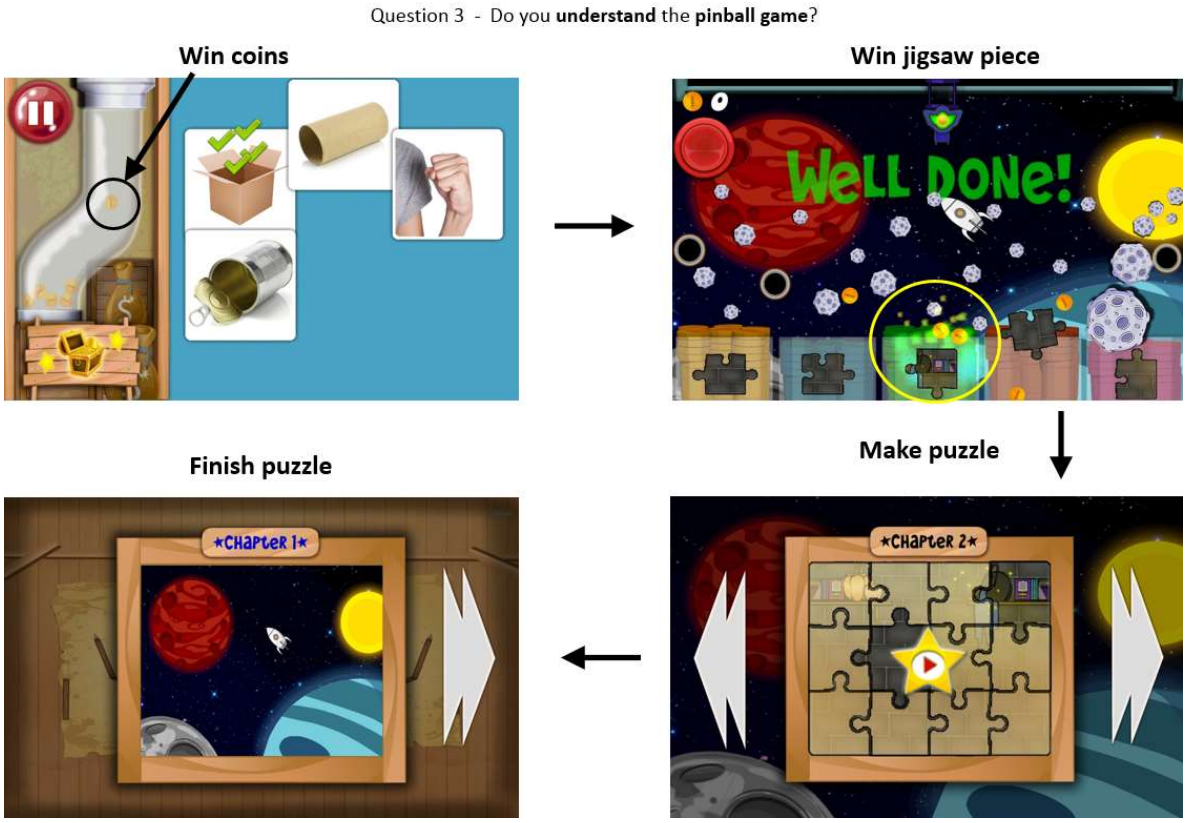


Figure 4-2 Example of visual aid to present concept visually

4.1.1.1 Data analysis: Thematic Analysis

There are a number of different approaches which can be used to analyse focus group data. In the present study, an inductive approach was sought which could generate themes directly from the data, as there was no prior theoretical framework. A thematic analysis approach was chosen, which identifies, analyses and reports patterns (themes) within the data (Braun & Clarke, 2006). This approach was applied to transcription data collected from video recordings of focus groups, and followed the steps laid out in Figure 4-3, based on guidelines set out in Braun and Clarke (2006).

The first phase involved familiarisation with the data during transcription process. A coding framework is then developed to code parts of the data in a systematic manner. Individual codes are assigned to words, phrases or chunks of transcription. In the present study, more than one researcher took part in coding the transcription. To ensure inter-rater reliability, a coding comparison was added to the analysis pipeline to ensure consistency amongst coders. Following coding, a node is developed which combines all of the individual codes into one collective section. These nodes are then analysed by the researchers, and overarching themes are constructed. The present study used these themes to report key findings, and make subsequent recommendations for the prototype of the application. (A detailed description of this approach is provided in Chapter 1 – Results).

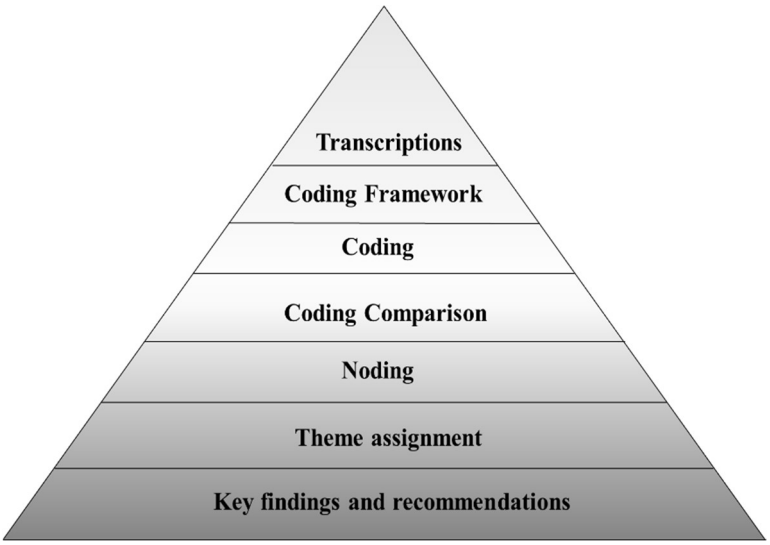


Figure 4-3 Thematic data analysis steps in the present study

In thematic analysis, prevalence of responses is typically indicated with terms such as "many participants" or "a number of participants" (Braun & Clarke, 2006). Although Braun and Clarke identify problems with these terms due to their vague nature, they have been used in the present study, in line with convention, to avoid quantifying numbers of participants, typically used in alternative approaches which focus on frequency of response. The aim of the present study was to explore the Listen-In prototype in a small group of participants, which was not intended to be representative of the heterogenous aphasia population. As such, it was not considered necessary to generate specific information on frequencies of responses, but rather bring wider issues to the forefront which needed to be addressed during development. (For example, it was not considered important if nine out of twelve participants preferred the colour red for a button, given the individual nature of preferences. Here, the key issue was whether the overall use of colour was positively regarded and that the button was clear and easy to use).

4.1.2 Study design

4.1.2.1 Ethics

Ethical approval for the Listen-In project was obtained from National Research Ethics Service Hampstead Committee, London (15/LO/0569).

4.1.2.2 Design

A timeline of development is shown in Figure 4-4. Five focus groups were held with the UK charity Connect over the period of approximately one year. This provided access to a venue with a ready-made source of recruitment as well as experienced facilitators and moderators who have worked with PWA. The research team chose not to fulfil the role of facilitators or moderators, as this can lead to bias in data collection. The researcher's role in these groups was in developing the questioning route prior to groups, and providing general support during each group, including data collection via field notes and video recordings.

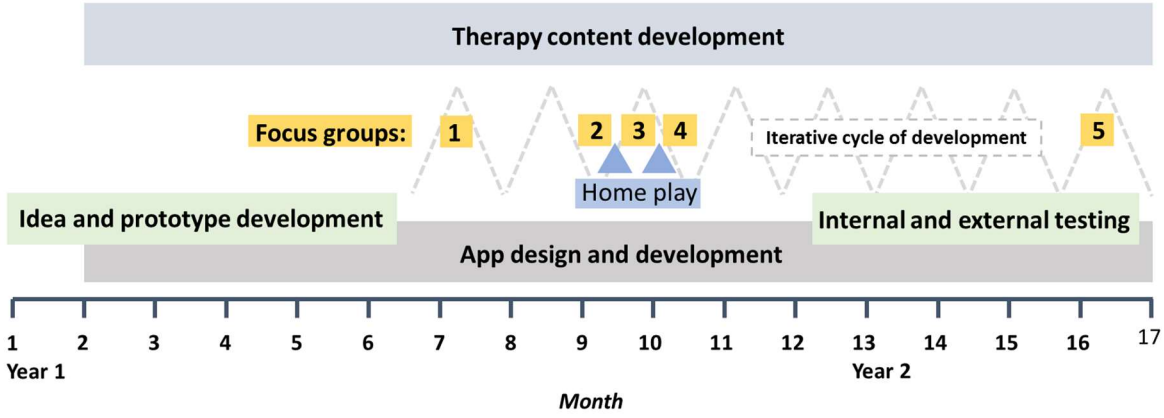


Figure 4-4 Timeline of Listen-In development

4.1.2.3 Recruitment

Participants were recruited via opportunity sampling at a weekly drop-in session at Connect, one week prior to the first focus group. I introduced the project by giving a short presentation, and invited interested members to take part. Connect staff also contacted several members by telephone. All participants were regular members of Connect and therefore most knew each other prior to the focus groups.

4.1.2.4 Inclusion and exclusion criteria

The main inclusion criterion was aphasia due to stroke. A broad criterion was used in order to obtain as wide a range of participants as possible for exploratory purposes. It was not

deemed necessary for participants to have a receptive language impairment, due to the nature of the general feedback that was required to meet the studies aims. Other inclusion criteria included being willing and able to attend one or more groups, consenting to be video-recorded, able to give informed consent (with support as required), and be over the age of 18. Exclusion criteria were presence of another medical condition or hearing or visual loss which would severely impact a person's ability to use a computer tablet.

4.1.2.5 Participants

Twelve participants took part over five focus groups (Table 4-2). Not all participants attended all focus groups. Individuals varied in age (M=58 years), time since stroke (M=7;01 years), aphasia profile, and previous experience with technology.

4.1.3 Listen-In: Speech comprehension therapy program

Listen-In is a speech comprehension therapy program consisting of spoken word to picture matching tasks with gamification. This section describes the therapeutic component of Listen-In as used by patients in the trial. For a description of the overall structure of the app, including gamification, see Chapter 1.

4.1.3.1 Spoken word-to-picture matching task

Listen-In is an app based spoken word-to-picture matching task. A participant hears a target word or sentence, and is required to choose a matching picture (Figure 4-5). Participants receive feedback in the form of visual ticks and crosses and respective sound effects. If the selection is wrong, the chosen picture is taken away, and the participant can listen again, and make another choice. This continues until the participant selects the correct picture, meaning the correct word-picture pairing is always selected. If the selection is correct, the participant moves on to the next challenge. A repeat button in the middle of the screen enables participants to repeat the spoken stimulus an unlimited number of times within each challenge. This task was designed to target phonological and

semantic processing, through inclusion of phonological and semantic foils (Goodglass & Kaplan, 1972). Completing this task would engage multiple levels of processing: according the cognitive neuropsychological model of single word processing (Whitworth et al., 2014), this would include multiple modules, from auditory input analysis, through to the semantic system. As such, patients with a range of speech comprehension impairments may benefit from different aspects of treatment, therefore this treatment can be considered a broad approach.

The design of the task can be viewed as an errorful approach with feedback. However, as patients are constrained to up to five different response, errors are restricted to these particular items. Errorful approaches have been demonstrated to support learning to a similar extent as error-reduction approaches (Mcclelland et al., 2002). The key therapeutic components are twofold: feedback to reinforce correct responses (reinforcement learning); and exposure to the correct word-picture pairing on each trial to promote paired associate learning (Hebbian learning).

Figure 4-5 Example challenge in Listen-In. The target word "ring" is flanked by a semantic distractor "finger", and phonological distractor "fin".

Partici-pant	Gender	Age	Time since stroke (Y;M)	Focus Group:					Speech production	Previous technology experience				
										Mobile phone		Tablet		SLT
				1	2	3	4	5		General	Games	General	Games	programs
P1	F	52	6;03	✓	✓	✓	✓	✓	Fluent, some sentences	✓	-	-	-	✓
P2	M	55	4;11	✓	✓	✓	✓	✓	Fluent, full sentences	✓	✓	-	-	✓
P3	M	70	7;10		✓	✓	✓		1-2 single words	-	-	✓	-	-
P4	M	50	8;00*		✓	✓	✓		Single words	✓	-	✓	-	✓
P5	M	42	0;08	✓	✓		✓		Fluent, full sentences	✓	-	✓	-	-
P6	M	65	1;01	✓				✓	Non-fluent, short sentences					
P7	M	57	11;09		✓			✓	No speech output	✓	-	✓	-	†
P8	M	44	2;09					✓	Non-fluent, short sentences					
P9	M	74	14;00*	✓	✓	✓			Fluent, full sentences	✓	-	-	-	-
P10	F	66	15;00*	✓	✓	✓		✓	1-2 single words	✓	✓	✓	✓	-
P11	M	61	4;08	✓					Non-fluent, short sentences					
P12	M			✓					Non-fluent, short sentences					
		M=58	M=7;01	8	8	6	5	6						

*Table 4-2 Characteristics of focus group participants. *Time since stroke rounded to nearest year. † Unknown. Shaded regions=information not available.*

4.1.3.2 Therapy stimuli

Listen-In contains a core set of treated words ('lexical items'), composed of all major word classes (Table 4-5). A corpus was used to identify the most frequent words in spoken English, and these were curated to form a final word list of approximately 870 items (van Heuven, Mandera, Keuleers, & Brysbaert, 2014).

Word type	Set A Therapy Challenges		Set B Therapy Challenges	
	Lexical items	Total challenges	Lexical items	Total challenges
Nouns	652	2285	657	2243
Verbs	132	329	137	333
Adjectives	70	144	72	146
Pronouns	6	198	6	198
Prepositions	13	289	13	289
Tense	(9 verbs)	54	(9 verbs)	54
Total	873	3299	885	3263

Table 4-3. Therapy content of Set A and Set B. Lexical items are number of unique lexical items (e.g. ball, girl, red). Challenge items are the total number of challenges created from these lexical items, and can be single words, phrases, or carrier sentences.

An individual challenge is made up of a target word, and up to five foils (Figure 4-5). Fifteen individual challenges form one 'block'. Each lexical item is trained in more than one challenge, so that the lexical item is presented singly, or within a phrase or a sentence context (linguistic construction). Table 4-3 displays the range of linguistic constructions. In this way, words are presented in linguistic contexts more akin to everyday comprehension, in addition to being presented as a single word. The number of challenges per lexical item varied. Each challenge is typically composed of two semantic foils (associate and/or coordinate), two phonological foils (minimal pair, or other close competitor), and one unrelated foil. It was not possible to find appropriate foils for all words; therefore, there is variation in the ratio of foil types across challenges. For example, if a close phonological distractor was not found, a distractor with one or more overlapping phonemes would be used.

In order to test trained and untrained items, a secondary outcome measure was created (Auditory Comprehension Test, ACT) with two matched sets (Set A, Set B, 220 items total). Items for this test were selected from core therapy set:

Listen-In therapy:

Core therapy items

+

220 ACT items:

- 110 trained items (trained in Listen-In, tested in ACT)
- 110 untrained items (not trained in Listen-In, tested in ACT)

As a result, two therapy sets were formed. Therapy Set A contained 110 Set A ACT items, in addition to the remaining core therapy items (items not used in the ACT). Therapy Set B contained 110 Set B items, plus core therapy items. There was a small difference in number of lexical items and challenges in therapy Set A and Set B (Table 4-6). Audio stimuli were produced and processed in a uniform manner by a professional audio recording company (https://www.soundcuts.net/). Speakers were all native southern British English speakers (three male, two female).

4.1.3.1 Adaptive algorithm

In collaboration with a further member of the team, I developed an algorithm which automated challenge presentation within the app, which was adaptive based on individual performance. Each challenge had a fixed, predetermined difficulty level (between 0.1 and 1, see below). At the beginning of therapy, all participants were presented with a pool of low difficulty items. Following completion of one block, the player was moved 'up' or 'down' in difficulty depending on performance. If the participant achieved >=70% accuracy over one block of 15 challenges, difficulty was increased; 40-70% difficulty remained the same; and <=40%, difficulty was decreased. The aim was to stop participants from receiving challenges at base or ceiling level. This is in line with the recommendation that treatment be presented just below an individual's maximum performance. Once the participant reached the maximum level of difficulty, the algorithm moved the player down to the middle level of

difficulty. Subsequently, players continued in this cycle until the completion of the therapy block (12-weeks) (Figure 4-7). Individual challenges for each block were selected at random from a pool of lexical items according to the level of difficulty required. Selection parameters were as follows:

Select all accurate challenges from the previous 50 blocks

Select challenge neighbours (30 challenges with difficulty value within .8)

Exclude neighbour challenges with exposure >100 times

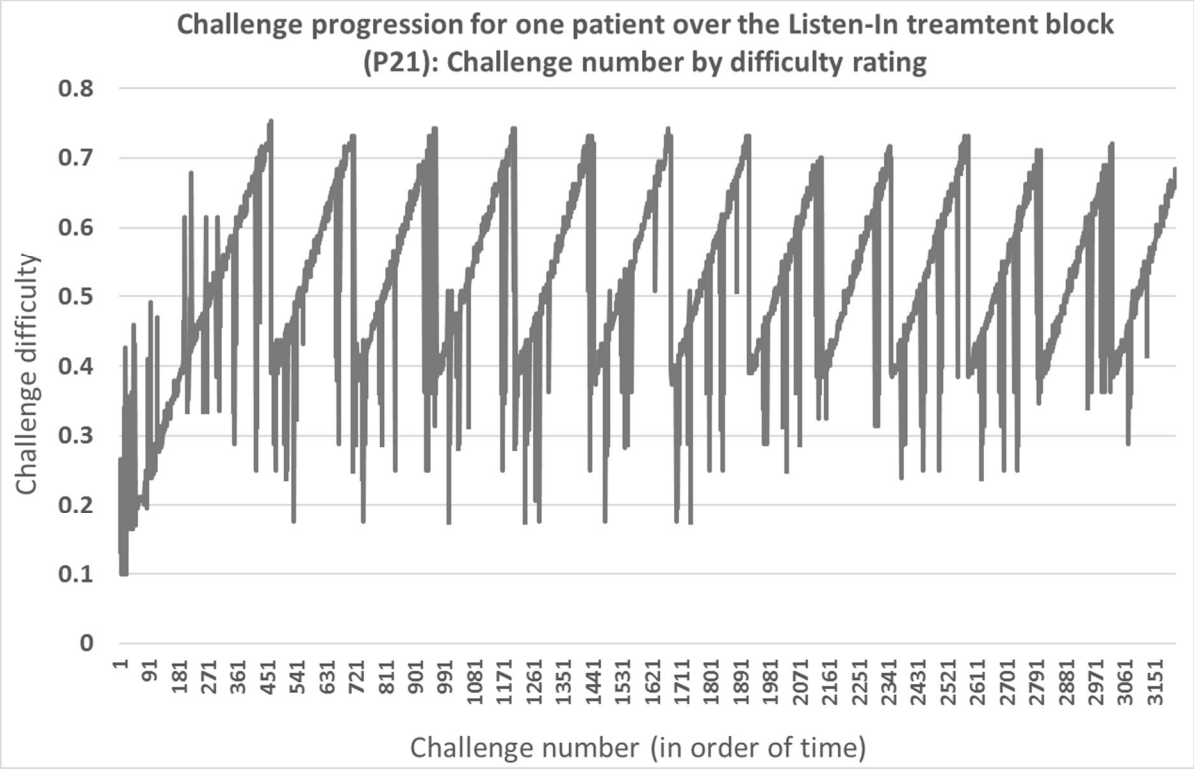


Figure 4-6 Progression through Listen-In challenges for one patient (P21) over time, by difficulty rating.

Word class	Linguistic context	Type of carrier sentence	Number of challenges		Example challenge (target word underlined)
			Set A	Set B	
Noun	Single word	1 syllable	271	259	Air
		2 syllable	268	267	Cobweb
		3 syllable+	90	89	Marmalade
	Phrase	1 syllable	269	268	A book
		2 syllable	268	268	A snowman
		3 syllable+	90	89	The company
	Sentence	Attributive	264	260	A bright parrot
		Predicative	281	278	The art is red
		Intransitive	52	54	The couple smile
		Intransitive + preposition (beginning)	62	51	The bull stands on the grass
		Intransitive + preposition (end)	40	34	The boy sits by the stairs
		Transitive (beginning)	69	63	The lady uses the phone
		Transitive (end)	174	167	He cleans the room
		Transitive + preposition	82	89	He gets a record for his birthday
Verb	Sentence	Intransitive	101	101	The boy peeks
		Intransitive + preposition	79	79	The child builds during playtime
		Transitive	81	82	He thieves some money
		Transitive + preposition	50	54	She tears the paper at work
Adjective	Sentence	Intransitive	10	12	The fat dog sniffs
		Phrase	72	72	The clean floor
		Sentence	62	62	The ball is wooden
Preposition	Phrase	Preposition phrase	145	145	Above the bag
	Sentence	Preposition sentence	144	144	The cat is behind the books
Pronoun	Phrase	Personal	126	126	She builds
		Possessive	36	36	Their dog
	Sentence	Possessive	36	36	Her ice-cream is delicious
Tense	Sentence	Progressive	27	27	He is cycling / he was cycling / he will be cycling
		Simple	27	27	She paints / she painted / she will paint
TOTAL			3299	3263	

Table 4-4 Linguistic constructions used to train single words within the Listen-In therapy program.

The algorithm produced a difficulty value for each challenge between 0.1 and 1 based on four parameters. Each parameter received an equal weighting:

Frequency = (((1 - 0.1) * (frequency - 1)) / (8 - 1)) + 0.1

Concreteness = (((1 - 0.1) * (concreteness - 1)) / (7 - 1)) + 0.1

Linguistic type = (((1 - 0.1) * (linguistic type - 0)) / (2 - 0)) + 0.1

Number of distractors = (((1 - 0.1) * (number of distractors - 2)) / (5 - 2)) + 0.1

Difficulty value = (frequency + concreteness + linguistic type + number of distractors) / 4

Details for the difficulty parameters are as follows:

(1) Two psycholinguistic variables: frequency and concreteness.
    a. Frequency refers to how often a word appears in a specific language. I used an up to date corpus of spoken English to obtain a frequency value for each target word (van Heuven et al., 2014), and then sorted these into frequency 'bins'. There is some evidence that low frequency words are harder to process for PWA, therefore low frequency items were given harder difficulty ratings (DeDe, 2012).
    b. Concreteness can be defined as "the degree to which a word's referents can be perceived through the senses" (Kiran, Sandberg, & Abbott, 2009). A corpus was used to obtain concreteness values for each lexical item (Brysbaert, Warriner, & Kuperman, 2014). If no value was available, a synonym was identified, and the concreteness of this word was used. Behavioural evidence suggests that PWA show better performance for highly concrete words (Kiran et al., 2009), therefore low concrete items were given harder difficulty ratings. These were also sorted into a set of concreteness 'bins'.
(2) Number of foils. Evidence has found larger array sizes are harder for PWA (Howland & Pierce, 2004). In this app, challenges with more foils were assigned greater difficulty ratings than those with fewer foils.

(3) Type of carrier sentence. Evidence shows PWA experience more difficulty comprehending words in a sentence context (DeDe, 2012). PWA also show within-subject instability in processing different syntactic structures (Caplan, Waters, DeDe, Michaud, & Reddy, 2007), and one theory is reduced processing capacity. A range of simple sentences were developed to provide a context for lexical items, and these were split into three categories, from easy to difficult: single words, medium, and hard. In this case, the assumption was that longer sentences would increase processing load, and therefore be more difficult for PWA.

Two further variables (not included in the above algorithm) were implemented at a block level (across 15 challenges) to manipulate the level of difficulty:

(4) Background noise. PWA frequently report difficulty understanding speech in noise (Rankin, Newton, Parker, & Bruce, 2014). Eight environmental sounds were used to introduce background noise and thereby increase level of difficulty. These noise files were: pub, forest, city/traffic, office, shopping centre, supermarket, playground, train station. Noise was introduced following initial completion of 50 blocks, and was subsequently presented on every fifth block. To manipulate difficulty further, the signal-to-noise ratio was increased according to performance-related criterion, and included the following ratios: -5dB, 0dB, 5dB, 10dB, 15dB. Performance related criterion was the difference in accuracy between the prior noise-free block, and the subsequent noise block, across five consecutive noise free/noise blocks (10 blocks in total). The difference in accuracy determined whether the SNR changed on the subsequent block:

Positive difference in accuracy = increase SNR ratio by one step

Negative difference in accuracy = decrease SNR ratio by one step

Zero difference in accuracy = SNR stays the same

(5) Acoustic distortion. PWA also report difficulties understanding speech in suboptimal conditions, such as on the telephone, and show susceptibility to this in single word comprehension tasks (Moineau, Dronkers, & Bates, 2005). To emulate this effect, a bandpass filter was applied to audio stimuli, which reduced the number of frequency bands, creating a reduction in acoustic information. This

filter attenuated speech stimuli under 500Hz and over 3000Hz to -26dB. Bandpass filtering was integrated as the first level in the background noise ladder: noise free, phone voice, -5dB, 0dB, 5dB, 10dB, 15dB

4.2 Chapters 2-4: Randomised controlled trial of Listen-In

4.2.1 Ethics

Ethical approval for the Listen-In project was obtained from National Research Ethics Service Hampstead Committee, London (15/LO/0569).

4.2.2 Design

A randomised, repeated measures cross-over design was used, with five, evenly-spaced testing time points (T1-T5) at 12-week intervals (Figure 4-8). Baseline language tests were administered at T1. The interval between T1 and T2 measured spontaneous change before therapy, and also controlled for test re-test effects, familiarity, and regression to the mean. The intervals between T2-T3 and T3-T4 formed therapy and standard care blocks, the temporal order of which was pseudorandomised across subjects. Maintenance timepoints were T4 and T5. Magnetic resonance imaging (MRI) scans were obtained for a subset of patients who were able to be scanned (N=25) at T2, T3 and T4. During the 12-week therapy block, patients self-administered Listen-In at home on a computer tablet, with a target of 100 hours of mass practice over 12 weeks (approximately 80 minutes per day). The standard care block formed a 12-week control block, and consisted of patients' usual daily activities, including any speech and language therapy unrelated to this study. Testing sessions were conducted either at the Institute of Cognitive Neuroscience, University College London, or in patients' homes.

4.2.3 Block randomisation

Participants were allocated into one of two cross-over arms of the study. The principle investigator, who was blinded to the identity of participants, performed the randomisation procedure. Participants were randomised to each arm sequentially using a minimisation method (Altman & Bland, 2005). This aimed to minimise differences between groups across three factors: Spoken Word and Spoken Sentence subtests on the Comprehensive Aphasia Test (CAT) (Swinburn, Howard, & Porter, 2004), and time since stroke. Allocation to a particular block was not possible for two participants due to prior

commitments; in these cases, they were forced into the required arm of the study, and minimisation carried on as usual for subsequent participants.

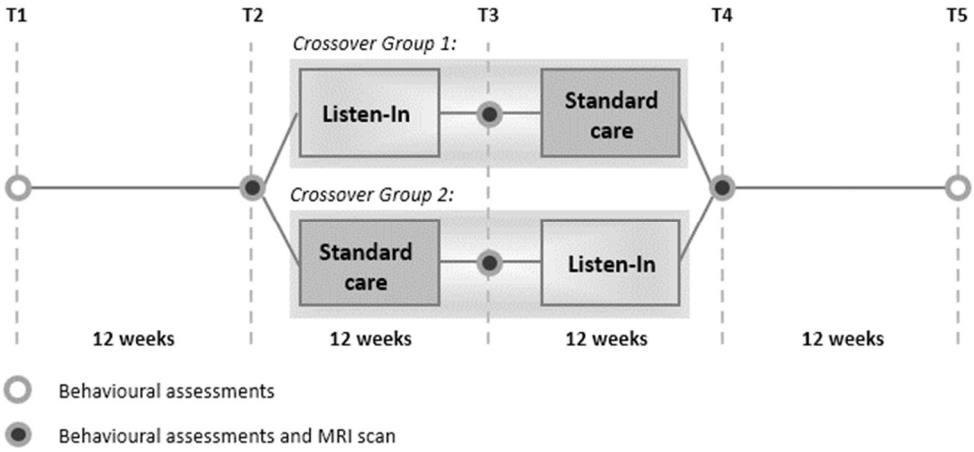


Figure 4-7. Flow chart of the randomised cross over study design

4.2.4 Blinding

It was not possible to blind researches or participants to their allocated treatment block, due to the nature of the intervention. All members of the team conducted assessments with participants at all time points.

4.2.5 Participants

Thirty-eight patients were recruited into the trial. Three patients data were excluded from analyses: one patient withdrew following T3 as they did not feel Listen-In was beneficial; one patient was withdrawn at T1 as their speech comprehension scores were found to fall outside the inclusion criteria; and one patient withdrew after T3 due to illness. Thirty-two patients completed the full trial (T1-T5) and three patients completed T1-T4, therefore data are reported in this thesis for thirty-five patients (Table 4-4). Recruitmen was from the PLORAS database (Wellcome Trust Centre for Neuroimaging, UCL; (Price et al., 2010)), a local outpatient aphasia clinic, and from focus groups carried out in Chapter 1 of this study.

Patients were recruited to reflect a typical clinical caseload and were therefore not excluded on the basis of hearing impairment, and as such patients had a range of hearing

levels within the normal to moderate-severe range (tested with pure-tone free field audiometry at 1000, 2000 and 4000 Hz) (see Table 4-4). Six patients wore hearing aids, and corrected hearing levels are reported. Two patients were classified as having severe hearing loss.

4.2.6 Inclusion and exclusion criteria

Inclusion criteria were: (i) more than six months post stroke; (ii) English as a dominant language; (iii) speech comprehension impairment at both single word and sentence level (aphasic cut off scores on the CAT: Spoken Words 25/30 or below; Spoken Sentences 27/30 or below). Exclusion criteria were: (i) a premorbid significant neurological (e.g. degenerative brain disease) or psychiatric disorder, (ii) not able to give informed consent. Patients gave written informed consent before commencing the study.

4.2.7 Behavioural assessments

4.2.7.1 Speech Comprehension Outcome Measures

4.2.7.1.1 Comprehensive Aphasia Test

An electronic version of the CAT was used as the primary outcome measure for speech comprehension. It is a widely standardised assessment battery used in clinical practice and research to obtain a broad overview of a persons' speech and language profile across multiple domains. Participants were administered the full CAT (minus the disability section) on a Windows laptop computer, at all time points. The primary purpose of this assessment was to capture change on speech comprehension subtests. These were Spoken Word and Spoken Sentence Comprehension. Spoken Word comprehension consists of 15 single nouns, and Spoken Sentence comprehension consists of 16 sentences. Sentences consist of different syntactic structures (e.g. reversible sentences) and test syntactic comprehension. In each subtest, the participant hears the item presented aurally over speakers, and chooses the matching picture from a choice of four. Two points are awarded for a timely correct response. One point is awarded if the response is self-corrected, repeated, or if the response is delayed (>5 seconds).

Table 4-5. Patient characteristics (N=35)

ID	Group	Sex	Handedness	Age at recruitment (years)	Time since stroke (months)	Type of stroke	Lesion volume (cm^3)	Free field Audiometry (dB) (1000, 2000, 4000Hz)	Hearing level
P1	1	M	Right	76	9	Haemorrhage	157*	63	Moderate**
P2	1	M	Left	65	88	Ischemic	274	81	Severe**
P3	1	F	Right	57	118	Haemorrhage	-	-3	Normal
P4	1	M	Right	36	21	Ischemic	116	1	Normal
P5	1	M	Right	71	52	Haemorrhage	227*	8	Normal
P6	1	M	Right	60	22	Ischemic	105	28	Mild
P7	1	M	Left	59	156	Ischemic	334	28	Mild
P8	1	F	Left	26	9	Haemorrhage	126*	-3	Normal
P9	1	M	Right	72	110	Haemorrhage	172	27	Mild
P10	1	M	Right	67	57	Ischemic	214	23	Mild**
P11	1	F	Right	76	44	Haemorrhage	194	28	Mild
P12	1	M	Right	47	89	Ischemic	221	13	Normal
P13	1	M	Right	63	270	Ischemic	293	12	Normal
P14	1	F	Right	74	6	Ischemic	-	12	Normal
P15	1	F	Right	61	113	(Unknown)	286	40	Mild
P16	1	M	Right	49	77	Ischemic	-	-	Not tested - none reported
P17	1	M	Right	50	16	(Unknown)	-	22	Mild
P18	1	M	Right	66	56	Ischemic	-	38	Mild

P19	2	M	Right	57	84	Ischemic	-	17	Normal**
P20	2	M	Right	52	109	Ischemic	246*	5	Normal
P21	2	M	Right	72	121	Ischemic	236*	13	Normal
P22	2	M	Left	82	111	Ischemic	171*	35	Mild
P23	2	M	Right	63	14	Ischemic	230*	20	Normal
P24	2	F	Right	70	127	(Unknown)	162	43	Moderate
P25	2	M	Right	80	92	(Unknown)	226	61*	Mod-severe sensorineural**
P26	2	M	Right	66	73	Ischemic	258	12	Normal**
P27	2	M	Right	46	25	Ischemic	182	2	Normal
P28	2	M	Right	59	66	Ischemic	232	60	Moderate
P29	2	M	Right	55	69	(Unknown)	296	7	Normal
P30	2	M	Right	47	40	Ischemic	-	37	Mild
P31	2	M	Right	48	32	Ischemic	-	35	Mild
P32	2	F	Right	74	222	Ischemic	179	-	Unable to complete test
P33	2	M	Left	60	115	Ischemic	252	37	Mild
P34	2	M	Right	63	14	(Unknown)	-	47	Moderate
P35	2	F	Right	50	39	Haemorrhage	157	-	Not tested - none reported
Mean (SD)				**61(12)**	**76(59)**		**216(59)**		
Group 1				60(14)	73(66)		209(73)		
Group 2				61(11)	80(53)		223(41)		

*Table 4-5 Patient characteristics on key demographic and baseline data. * 1.5T strength scanner (all other patients 3T). **Patient uses hearing aids, hearing levels are reported for corrected hearing*

4.2.7.1.2 Auditory Comprehension Test

The Auditory Comprehension Test (ACT) was used as the co-primary outcome measure to capture change in participants' speech comprehension for treated and untreated words. The ACT was developed to meet two key requirements: inclusion of a large number of items to be sensitive to change, and incorporation of trained and untrained items, to assess whether treatment effects are item-specific, or generalise to untreated items.

4.2.7.1.2.1 Content

The ACT consisted of 220 challenges, divided into Set A and Set B. Challenges were drawn from the therapy content, as described previously. Target words consisted of nouns, verbs and adjectives. The ratio of these word types reflected the ratio of word types in Listen-In therapy. Each lexical item was matched to create a matched set of 110 pairs across Sets A and B. Matching was based on word class, type of carrier phrase/sentence, frequency, concreteness, and syllable length. Each challenge consisted of one target item, two phonological foils, two semantic foils, and one unrelated foil. The unrelated foil was semantically related to a phonological foil.

4.2.7.1.2.2 Trained and untrained items (Set A and Set B)

Participants were randomised to receive either Set A or Set B therapy sets. Participants were tested on all 220 ACT items at all time points. Male and female recordings were randomly selected for each challenge. All pictures in the ACT were different from pictures used in the Listen-In therapy, to avoid the confound of identity priming.

4.2.7.1.2.3 Presentation

The test was presented on Windows 10 touchscreen laptop, using Matlab (version 15a). Challenges were presented in 8 blocks of 31 items, to allow for short breaks if needed, and order of presentation was randomised. Participants scored 1 point for a correct answer, and 0 points for an incorrect answer. Participants were allowed to repeat the target once (i.e. hear the target a maximum of two times) before responding, but this did not affect scoring.

4.2.7.1.2.4 Pilot testing

The ACT was pilot tested with a group of PWA (N=7) and a group of healthy age matched controls (N=22). To assess concurrent validity, PWA also completed CAT-words and CAT-sentences. Mean performance of the healthy control group was 95% (SD=2) indicating that healthy individuals were able to accurately comprehend and match the majority of stimuli. Mean performance of PWA was 60% (SD=15). This range suggests patients are able to understand some of the test items and that the test is sensitive to different levels of impairment. The ACT showed high concurrent validity with CAT-words (r=.75, p=.03, one-tailed). This suggests the ACT is likely to be a measure of the same underlying construct as CAT-words, and supports its validity as a test of spoken word comprehension. No correlation was found between the ACT and CAT-sentences (r=.33, p=.47, two-tailed), in line with the different underlying constructs of these two measures (syntactic versus word comprehension).

To investigate if there were difference in performance between Set A and Set B, accuracy was compared for PWA and healthy controls. PWA showed significantly worse performance on Set A (M=60, SD=16) compared to Set B (M=69, SD=16) (t(6)=-6.11, p=.001). For healthy controls, there was a small but non-significant difference between Set A (M=104, SD=3) and B (M=105, SD=3) (t(21)=-2.00, p=.06). These findings suggest a small difference in difficulty between sets. However, as this difference is small, no further changes were made.

4.2.7.2 Cognitive and Linguistic Assessments

Further cognitive and linguistic assessments can be seen in Table 4-9. These consisted of three baseline assessments, and ten repeated measures assessments.

Assessment	Abbreviation	Domain	Time-point
Auditory processing			
Auditory Comprehension Test	ACT	Spoken words	T1-T5
Spoken Words*	CAT-words	Spoken words	T1-T5
Spoken Sentences*	CAT-sentences	Spoken sentences	T1-T5
British Picture Vocabulary Scale (2nd ed.)	BPVS	Spoken words	T1-T5
Phoneme Discrimination Test	PDT	Phoneme discrimination	T1-T5
Environmental Sounds Test	ENVASA	Environmental sound identification in noise	T1-T5
Cognitive and Linguistic			
Naming*	CAT-naming	Naming	T1-T5
Repetition*	CAT-repetition	Repetition	T1-T5
Sustained Attention to Response Task	SART	Sustained attention	T1-T5
Digit span forwards*	Digit-F	Verbal short-term memory	T1
Culture Fair Intelligence Test Scale 2: Tests 1 & 2	CATTELL	Non-verbal fluid intelligence	T1
Semantic Association Test	SAT	Semantic system	T1
Self-report			
Self-report questionnaire		Speech comprehension and production in everyday life	T1-T5

*Table 4-6 Baseline and repeated measures assessments. *Subtests from the Comprehensive Aphasia Test.*

4.2.7.2.1 Baseline behavioural assessments (T1)

4.2.7.2.1.1 Fluid Intelligence

A measure of patients' fluid intelligence was taken at baseline to investigate how performance within this domain may relate to treatment outcomes. The Culture Fair Intelligence Test (Scale 2: Tests 1 and 2) (Cattell, 1940) was administered due to its minimal verbal requirements. Performance was calculated as a composite score from Tests 1 and 2, and converted into a percentage. Test 1 (12 items): Complete a sequence of four drawings by selecting the correct picture from a choice of 5. Test 2 (14 items): Pick one out of five drawings which is different from the others.

4.2.7.2.1.2 Semantic processing of objects

The picture-only version of the Semantic Association Test was used as a measure of patients stored semantic knowledge (Visch-brink, Stronks, & Denes, 2005). Patients were required to choose the picture, from a choice of four, which was semantically associated with the picture in the centre. There were 30 test items, with a maximum score of 1 for each item. There are no published norms for this assessment. In an abridged version with 10 items, a group of 27 healthy controls scored at ceiling (9.8 out of 10, SD=.40). In contrast, a group of 195 chronic aphasic patients made, on average, two errors, and showed greater inter-subject variability in performance (8.4 out of 10, SD=2.03) (Swinburn et al., 2004).

4.2.7.2.2 Repeated measures behavioural assessments (T1-T5)

Three tests were selected to measure different aspects of auditory processing: environmental sound discrimination, phoneme discrimination, and spoken word comprehension. One test investigated patients' sustained attention and was used as a non-language control measure. The final test investigated functional changes in patients self-reported everyday language use.

4.2.7.2.2.1 Environmental Sound Test

Aphasic individuals demonstrate impairments in both verbal and non-verbal auditory domains. For this reason, a measure of non-verbal auditory processing was incorporated into the test battery to assess the specificity of any treatment effects (I.E. if treatment gains are specific to verbal and/or non-verbal auditory domains). The Environmental Sound Test (ENVASA) was delivered on a laptop computer, and measured patients' ability to identify environmental sounds in natural auditory scenes. The test was adapted from Leech, Gygi, Aydelott and Dick (2009), and contained 80 trials.

In each trial, the target picture and sound were presented simultaneously (e.g. cat and 'meow'), followed by pause of .5 seconds, and then a stream of environmental noise for 7 seconds. The patient pressed a button as soon as they heard the target sound. There were three variables: onset time, signal-to-noise ratio, and congruency. Onset time was the time interval between the environmental noise and the target sound, and was early (<1 second), or late (>1 seconds). Signal-to-noise was the difference in intensity of the signal relative to the noise, and was categorised as high (SNR=+3dB) or low (SNR=-6dB). Congruency was whether the background noise was contextually congruent (50% of trials) or incongruent with the target sound. For the purpose of the present study, trials were collapsed across conditions to produce an overall score, which was the number of trials which the participant correctly identified (I.E. responded within 2 seconds). Scores were converted to percentage correct.

4.2.7.2.2.2 Phonemic Discrimination Test

Individuals with aphasia show impairments in acoustic-phonological processing, and these impairments have been associated with speech comprehension deficits (Robson, Keidel, et al., 2012). The Phoneme Discrimination Test was developed to measure changes in this domain, and was based on the discrimination assessment used in Robson, Keidel, Lambon Ralph and Sage (2012). It is a bespoke, tablet-based assessment which measures patients' ability to discriminate consonant-vowel-consonant (CVC) non word strings. On each trial, the patient listens to three CVC strings, and presses a button which indicates the string which is different. On 36 trials, the initial phoneme in the string differed in either place, voicing or manner (e.g. zuth-zuth-zug**)**, and on 6 trials, the vowel was different (e.g.

fom-fom-fim). Correct answers were always in the first or last position in the string (AXB design). Raw accuracy across trials (N=42) was collapsed to produce a percentage correct score.

4.2.7.2.2.3 Spoken Vocabulary Comprehension

The British Picture Vocabulary Scale is a graded test of vocabulary comprehension designed for use in children, and provided additional information about patients' receptive vocabulary (Dunn, Dunn, Whetton, & Burley, 1997). This test was administered to provide further information about patients single word comprehension. In each trial, the target word was read aloud by the researcher, and the patients' pointed to the correct picture in the stimulus booklet from a choice of four. All patients began on Set 9, and the test continued until the stopping criteria were reached. A raw score was calculated according to the manual, and converted to a percentage score based on the maximum possible score (168).

4.2.7.2.2.4 Sustained Attention to Response Task

A non-verbal version of the Sustained Attention to Response Task (SART) was included as a non-linguistic cognitive measure (Manly, Davison, Heutink, Galloway, & Robertson, 2000). It was not expected to change over time, and was therefore used as a control measure to investigate specificity of treatment effects. The SART is a GO/NO-GO test delivered on a computer. Patients were required to press a button when they saw a boy appear on the screen (GO), and withhold a response when they saw a different boy appear (NO-GO). GO trials (N=191) and NO-GO trials (N=24) were presented in a pseudorandomised order. As NO-GO trials were infrequent and unpredictable, to be accurate, the patient was required to maintain attention across trials, and successfully inhibit a response to NO-GO trials. Total error score (SART-errors) was the number of incorrect 'hits' on NO-GO trials, and omissions on GO trials.

4.2.7.2.2.5 Self-report measure

A bespoke self-report task was developed which asked patients to rate how easy they found speech and language related activities in the last month. Three subsections related

to: (i) speech comprehension; (ii) speech production; and, (iii) general activities of daily life (control measure) (Appendix 2). The task was presented on Powerpoint, with one question on each slide, and a Likert scale underneath ranging from 'very easy' to 'very difficult'. The researcher read the question aloud (using total communication support where necessary), and the patient gave their response by pointing to the scale. Score was the number of points in each subsection, with greater points reflecting poorer self-ratings.

4.2.8 Behavioural Data Analyses

4.2.8.1 Group level analyses

Treatment effects

Change scores were calculated for each block (treatment, standard care) by subtracting scores over relevant blocks. For Group 1 this was T3-T2 for treatment, and T4-T3 for standard care; for Group 2 this was T4-T3 for treatment, and T3-T2 for standard care. Changes scores were analysed using repeated measures analysis of variance (ANOVA) for each repeated measures outcome measure. Model details are reported in Chapter 2.

Baseline stability

Baseline stability over T1 and T2, for selected assessments, was investigated using paired samples t-tests.

Maintenance effects

For assessments demonstrating significant treatment effects, investigation of maintenance of treatment effects were also planned. Change scores were calculated over maintenance blocks. For Group 1, this was T4-T3 (12 weeks post treatment) and T5-T4 (24 weeks), and for Group 2, this was T5-T4 (12 weeks). Change scores were analysed using paired samples t-tests.

4.2.8.1 Individual treatment effects

To investigate individual treatment effects for treated and untreated performance on the ACT, a McNemar test was selected. McNemar is a statistic test for paired nominal data, and as such is appropriate for correct/incorrect response type assessments. A 2x2 contingency table was used to test whether the frequencies between correct and incorrect response were different from pre to post treatment for individual patients. The null hypothesis is that there is no change in frequency of correct and incorrect response from pre to post treatment, and the alternative hypothesis is that there is a significant difference. The test statistic was calculated in Excel as follows:

A = Number of incorrect-to-correct responses (e.g. 28)

B = Number of correct-to-incorrect responses (e.g. 14)

X^2 = ((A-B)–1)^2 / (A+B)

X^2 = ((28-14)–1)^2 / (28+14)

X^2 = ((14–1)^2 / (42)

X^2 = 169 / 42

X^2 = 4.02

$X^2_{(.05, 1)}$ = 3.841

In this example for one patient (P1), the test statistic (4.02) is greater than the critical value, indicating significantly more correct-to-incorrect responses.

4.2.8.1 Predicting response to treatment

Simple correlations were planned to investigate associations between key demographic and behavioural variables, and treatment outcomes.

4.2.8.2 Explanatory modelling (automatic linear modelling)

Due to the large number of baseline measures collected, explanatory modelling was chosen to investigate which baseline variable, or combination of baseline variables, best explained response to treatment in the current sample of patients. Automatic linear modelling (ALM) in SPSS24 was used, which consists of linear regression with multiple predictor variables (Field, 2013). The advantage of this method compared to traditional

linear regression (such as step wise linear regression) is that it is possible to conduct a more comprehensive analysis by including all possible combinations of the independent variables, thereby considering all possible regression models. The all possible subtests method produces up to 10 best models, in order of how much variance they explain in the dependent variable. Statistical and theoretical information can then be used to decide on a final model. A further advantage of this method is automated data preparation. This includes replacing missing values, transforming categorical variables, and identifying influential outliers, termed Cook's Distance. This measure is an indication of the impact of the outlier on the model parameters, and is calculated for independent and dependent variables which are more than three standard deviations away from the mean. Values close to 1 indicate the variable has greater influence on the model, and inclusion or exclusion of these variables should be considered (Field, 2013). When using ALM for conducting a single model, without model comparison, the adjusted R^2 parameter can be selected. This parameter provides an estimation of how much variance in the dependent variable is explained by the model.

4.2.9 Magnetic resonance imaging (MRI)

4.2.9.1 Principles of MRI

Magnetic resonance imaging (MRI) is a non-invasive imaging technique which produces high quality images of internal anatomy. MRI exploits the principles of nuclear magnetic resonance (NMR): when atomic nuclei are exposed to a strong magnetic field, they absorb and reemit electromagnetic waves at a characteristic frequency ('resonant frequency') within the radio frequency range (RF) (Storey, 2006). The MRI scanner captures this signal and converts it into a high resolution image. MRI provides information about the macroscopic volume, or density, of brain matter; however, it does not provide a direct measure of underlying microscopic morphology of the brain. The following section describes the principles of MRI as applied to imaging of the human brain, and therefore focuses on hydrogen atoms.

4.2.9.2 Hydrogen nuclei in the magnetic field

Hydrogen atoms contain one proton in the nucleus and no electrons. All isotopes which have unpaired protons or neutrons contain a property known as nuclear spin, which is the rotation of the nucleus around its own central axis (Storey, 2006). The unpaired proton in the nucleus of the hydrogen atom creates a net spin characteristic, and produces a small loop of current. This circular loop behaves in the same way as a magnet, and produces its own small magnetic field, which is susceptible to externally applied magnetic fields. This force is known as torque, and can be visualised as a turning force on the nucleus. The magnetic field created by nuclear spin creates a further magnetic field perpendicular to it. This causes the nucleus to undergo its own magnetic moment, which is likewise affected by external magnetic fields, this time forcing the nucleus into alignment with the direction of the external field. In summary, hydrogen atoms possess small magnetic moments, which can be likened to a magnet. The magnetic moment of a nuclei is characterised by the relationship between the magnetic moment of the nucleus (μ), and its intrinsic spin (I). The gyromagnetic ratio (γ) is a constant property of the nucleus, and varies depending on the type of isotope. This relationship is expressed by the equation:

$$\mu = \gamma I$$

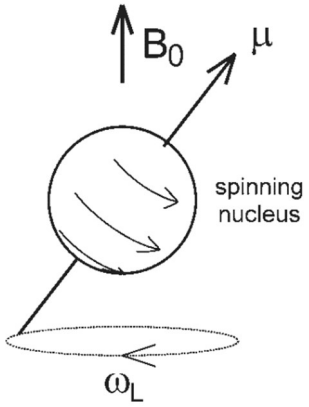


Figure 4-8 Precession of a nuclei in the presence of an external magnetic field (taken from Storey, 2006). Image reproduced with permission of the rights holder, Springer Nature.

When an external magnetic field is introduced (B_0), the magnetic moments rotate into alignment, in a process called precession (Figure 4-10). This alignment does not occur

immediately, but creates a 'wobble', or circular motion, around the vertical axis of the direction of the external field (B_0) (Storey, 2006). The frequency of this spinning rate is called the Larmor frequency (ω_L), and is proportional to the strength of the local magnetic field:

$$\omega_L = \gamma B_0$$

4.2.9.3 Magnetic resonance signal

The precession of the nucleus and the magnetic moment of the spin result in oscillating magnetic fields. These individual oscillations combine to form a net magnetic field, and this forms the basis of the MR signal (Young, 1984). In their natural state, the hydrogen ions will be precessing at random, and so the net signal will be zero (equilibrium). In the presence of a strong magnetic field, nuclei will begin to align, in synchrony, to be either parallel or anti-parallel with the direction of the magnetic field, producing a rotating magnetic field. In the MRI scanner, pulses of RF energy are emitted from transmitter coils. The frequency of the RF energy is tailored to match the Larmor frequency of the hydrogen atoms (I.E. their resonant frequency), so that it is absorbed by the nuclei. As they absorb this energy, the amplitude of transverse magnetisation gradually increases, and this signal is captured by RF receiver coils. The signal does not last indefinitely, as the nuclei eventually relax back to their usual state. The rate of relaxation varies according to tissue type, and these signal differences are exploited to enable differentiation of different tissues within the MR image.

4.2.9.1 Signal decay

Relaxation occurs in two ways. Longitudinal relaxation (T_1 recovery) refers to loss of energy from the nuclei due to their molecular environment (Storey, 2006). Fluctuations in the local magnetic field occur due to molecular tumbling with neighbouring electrons, nuclei and molecules. If the fluctuations are close to the Lamour frequency, then energy release occurs quickly, until equilibrium in the surrounding environment is reached (such as in lipids). If the fluctuations are higher or lower than the Lamour frequency, then the

rate of energy release will be slower (such as in cerebrospinal fluid). Repetition time (TR) between RF pulses in the scanner can be tailored according to these different longitudinal relaxation rates, to be sensitive to different types of tissue, known as 'T_1 weighting'.

Transverse relaxation (T_2 decay) refers to a loss of phrase coherence amongst the nuclear spins, and subsequent loss of net magnetisation (Young, 1984). Following RF excitation, the nuclei continue to spin in synchrony for a short time, but then relax back to their usual random state of precession, causing the net transverse magnetisation to gradually return to its equilibrium. Loss of spin coherence occurs due to inhomogeneities in the local and external magnetic fields. Variations in the local molecular environment affect the precession of spins and cause them to dephase (T_2 decay), due to neighbouring nuclei and molecules which possess their own small magnetic fields. T_2 decay occurs at different rates for different tissue (cerebrospinal fluid > grey matter > white matter) due to different rates of molecular tumbling. Variations in the external magnetic field of the scanner also influence relaxation (T_2*), referred to as free induction decay. This arises from both variations in the magnetic field of the scanner, and the magnetic susceptibility of different biological materials. Because of these different susceptibilities, spin dephasing is affected at the boundaries of materials, such as the sinuses (Storey, 2006).

During free induction decay (T_2*), there is an exponential loss of transverse magnetisation due to dephasing. This effect is faster than T2 decay, but is also constant and predictable, and can therefore be corrected for in a process known as spin-echo sequencing. During precession, spins gradually begin to dephase. A refocusing RF pulse is applied at 180 degrees, flipping the spins so they begin to swing back into alignment. The time this takes to occur is termed echo time (TE), and varies across tissue types. After this, the spins will again begin to dephase in the reverse direction, generating energy, known as spin echo. The characteristics of transverse relaxation (T_2 and T_2*) for different tissue types can be exploited in the same way as longitudinal relaxation (T_1) to produce contrast between tissues on MR images. In T_1 weighted images, the TE and TR are shorter to reduce the signal due to T_2 decay, and maximise signal from T_1 relaxation. Conversely, in T_2 weighted images, lengthening these parameters maximises the signal from T_2 decay.

4.2.9.2 The MRI scanner

The main component of the MR scanner is a primary magnet, which produces the external magnetic field (B_0) (Horowitz, 1995). The strength of the magnet is measured in tesla (T), which is typically 1.5T or 3.0T. The magnet is formed of coils of niobium-titanium wire encased within liquid helium, to produce the right temperature for superconduction . The patient passes through the middle of the coil where the field is most homogenous. An important aspect of the scanner is homogeneity of the magnetic field (B_0). As previously outlined, inhomogeneities can cause spin dephasing and disrupt accurate imaging. To compensate for this, shim coils emit compensatory magnetic fields, and gradient and higher order shim coils correct for linear and quadratic variations respectively.

4.2.9.3 Signal localisation

Three gradient coils produce magnetic fields with linear variations in strength, along three orthogonal directions: X, Y and Z . The purpose of these coils is to enable source localisation of the RF signal. During precession, only energy at the Lamour frequency (the resonant frequency) is absorbed and re-emitted by protons, and it is this signal which is detected by the RF receiver coils. All excited protons within a tissue emit this source simultaneously. To locate the source of the signal, excitation is therefore delivered selectively, in slices. Gradient coils achieve this using magnetic field gradients, whilst spatially encoding the phase and frequency of the signal. As the precession of protons is directly proportional to the strength of the external magnetic field (previously described in the Lamour equation), manipulating this field enables only certain protons to absorb and re-remit energy and be detected by RF receiver coils. RF transmitter coils produce the B1(t) magnetic field during RF excitation, and receiver coils detect the signal from the tissues during excitation.

The Z coil has positive and negative poles at each end of the bore, and sums with the field from the primary magnet (B_0) to create a graded magnetic field along the z-axis. Only protons which lie at the particular location on the z-axis where the Larmour frequency (ω_L) equals the RF field (ω_{RF}) will be excited (see Figure 4-11). In this way, the gradient

coils can control which protons emit energy. The amplitude of the gradients, along with the bandwidth of the RF pulse, can be manipulated to alter the slice thickness.

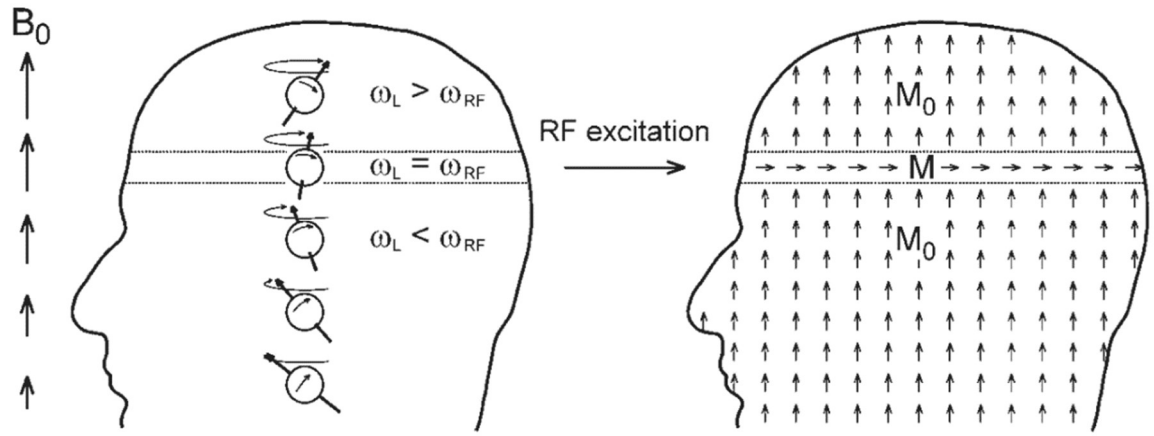


Figure 4-9 Radio frequency slice excitation (taken from Storey, 2006). Image reproduced with permission of the rights holder, Springer Nature.

To enable spatial localisation within a slice, gradient fields vary the precession frequency in the x-axis, and phase in the y-axis. In frequency encoding, the gradient is applied during RF signal acquisition. Tissue located in stronger areas along the gradient of the magnetic field will emit a higher frequency signal, meaning that variations in frequency will correspond to spatial locations. In phase encoding, a brief gradient pulse it emitted before the signal acquisition. This causes variation in spin phases. Along this axis, the signal from different phases therefore corresponds to spatial locations along the gradient. For phase encoding, multiple repetitions are needed with varying amplitudes, to extract spatial information.

Multiple samples are taken over time for each slice excitation. The enables information from phase encoding, which uses multiple repetitions, to be combined. The information is represented in a 2D array called k-space, and is visualised as a series of lines. Fourier transformation is applied to this data, which enables the spatial information from the signal to be reconstituted. The final product is a single map, which represents amplitude at each voxel.

4.2.9.4 The MR image

In T_1 weighted imaging, each voxel typically has a spatial resolution of $1mm^3$, and represents thousands of neurons, and other supporting structures, such as glial cells, at a macroscopic level. Grey matter is composed of neuronal cell bodies, glial cells, dendrites, blood vessels, extracellular space, and myelinated and unmyelinated axons, and is located in the outer 3mm of the cerebrum, whilst white matter is composed mainly of myelinated axons, glial cells, and extracellular space (Mills & Tamnes, 2014). The present study utilises T_1 contrast, which is sensitive to variations in myelin content between grey and white matter, therefore change in image intensity over time may reflect changes in myelination (Tardif et al., 2016). However, due to the indirect nature of MR imaging, underlying microstructure cannot be directly inferred. In the same way, the neurobiological processes underlying regional morphological changes observed in longitudinal imaging cannot be inferred using MRI. However, they have been proposed to include synaptogenesis, spine formation/elimination, dendritic branching/pruning, myelination, angiogenesis, and gliogenesis (Tardif et al., 2016).

4.2.10 Voxel Based Morphometry (VBM)

4.2.10.1 Principles of VBM

Voxel-base morphometry (VBM) is an automated technique which uses T_1 weighted MR images to investigate differences in volume across the whole brain, on a voxel-by-voxel basis (Ashburner & Friston, 2000; Mechelli, Price, Friston, & Ashburner, 2005). The aim is to identify regions of grey and white matter where variation in structure is associated with a variable of interest (e.g. auditory short term memory (Leff et al., 2009). A series of processing steps reduce the impact of large scale differences in gross anatomy, whilst remaining sensitive to small scale differences in the composition of local tissue. As the whole brain is investigated, rather than pre-selection of a particular structure, it is considered an objective and un-biased technique.

VBM uses segmented brain tissue images to identify differences between groups of individuals, or across individuals over time. Statistical Parametric Mapping (SPM12) is a commonly used software program which segments T_1 weighted volumetric images into GM, WM and cerebrospinal fluid based on tissue probability maps (Ashburner & Friston, 2005). Each voxel in the segmented image contains a value which indexes the probability that it belongs to a particular tissue class, and these values form the basis of VBM analyses. It is not clear how probability values relate to tissue composition, beyond providing volumetric information. Eriksson and colleagues (2009) investigated whether GM probability values, derived from MRI images processed using two different software programs (SPM5 and FSL-FAST), correlated with histological measures of GM, taken from epilepsy patients undergoing temporal lobe resection. The authors found that none of the neuropathological measures correlated with GM probability values, suggesting that probability maps used in VBM are not sensitive to variations in tissue composition, at least for the measures tested. Nevertheless, decreases in concentration or volume are commonly interpreted as a loss of GM tissue due to atrophy or neuronal cell loss (Seghier, Ramsden, Lim, Leff, & Price, 2014), and increases have been attributed to thickened GM (Hervais-Adelman, Moser-Mercer, Murray, & Golestani, 2017b), or difficulty with GM and WM demarcation due to pathological brain tissue. However, inferences regarding the precise nature of structural differences, or changes, remain speculative.

Three pre-processing steps are routinely implemented in VBM analyses: (1) Scans are spatially normalised into a common stereotactic space, such as Montreal Neurological Institute (MNI) space (as used in this thesis); (2) Normalised images are segmented into grey and white matter; (3) Normalised grey and white matter images are smoothed. Following pre-processing, a general linear model is used which relates the concentration or volume of grey or white matter to a particular variable of interest, resulting a statistical parametric map. This map represents a statistical test at each voxel. Due to the large number of voxels, correction for multiple comparisons is applied to reduce the rate of false positives (Type 1 error), using Guassian random field theory. This is based on the assumption that the spatial distribution of these statistics (I.E. voxel values) reflects a smoothly varying field, and calculates the likelihood of a voxel, or cluster of voxels, appearing by chance within this smooth field. The level of correction required depends

primarily on the smoothness of the images. The number of resels or resolution elements is calculated from the 3D statistical map using the Euler characteristic. In SPM, there are several different ways of correcting for multiple comparisons. I used best practice which is to report significant peaks or clusters at the Family Wise Error (FWE) corrected p-threshold. When I had no a priori expectations of where effects may be seen, this was corrected for the whole brain volume. Where I had anatomically constrained a priori hypotheses based on previous work, I used a small volume correction using a mask that covered the regions of interest.

4.2.10.2 Chapter 3: Lesion identification

Lesions consist of abnormal tissue composition, characterised by abnormal local magnetic properties. In VBM, these can lead to ambiguous voxel intensity values, and in some cases, intensity values may overlap with values from healthy grey matter, leading to misidentification during segmentation (Fiez, Damasio, & Grabowski, 2000). Abnormal intensity values can also have detrimental effects on spatial normalisation, leading to misregistration into stereotactic space. As such, VBM is not optimised for detecting focal lesions, and lesions may significantly disrupt the quality of preprocessing steps carried out.

Traditionally, the boundaries of lesions have been manually traced by trained individuals. However, manual tracing entails a degree of human decision-making, and this can introduce bias into results (Ashburner & Friston, 2000). An alternative method is to use an automated computer algorithm, which reduces the risk of bias, and objectively applies the same process to all scans. In this thesis, the Automated Lesion Identification (ALI) toolbox in SPM12 was used to identify patient lesions, and to carry out preprocessing steps prior to statistical analyses (Seghier, Ramlackhansingh, Crinion, Leff, & Price, 2008). ALI utilises optimised normalisation, and a modified segmentation technique, based on individuals with brain damage. This is based on two underlying assumptions; firstly, that voxels with atypical intensities will be outliers amongst grey and white matter segments; and secondly, that outlier voxels will have extreme intensities that are not within a normal

range of typical voxel intensities in healthy controls. Outlier voxels are then classified as an extra tissue class, allowing for the lesion to be identified.

Spatially normalised binary lesion data were obtained from patients' baseline T1 scans using Automated Lesion Identification in SPM12, with default parameters (Seghier et al., 2008). Binary images were checked against patient scans by hand, and regions incorrectly identified as primary lesion sites were manually removed from the binary image using MRIcron (Rorden & Brett, 2000). This resulted in one binary lesion image for each participant, where each voxel was classed as lesioned (1) or non-lesioned (0). Binary images were combined to produce a group lesion overlap map showing the distribution of patient lesions, displayed in Chapter 3.

4.2.10.3 Chapter 3: Cross-sectional VBM

Patients were scanned three times, twelve weeks apart, with some within and across patient variability in days between scans due to scheduling constraints. T_1-weighted whole brain images were acquired on either a Siemens 1.5T Avanto or a Siemens 3T Trio MRI scanner (Siemens Medical Systems, Erlangen, Germany), with a standard 32-channel head coil, using the same sequence. Patients were always scanned on the same strength scanner. A T_1-weighted 3D modified driven equilibrium Fourier transform sequence was used, which produced 176 contiguous sagittal slices with a 256 x 224 matrix, and resolution of $1mm^3$: repetition time/echo time/inversion time for 1.5T: 12.24/3.56/530ms, and for 3T: 7.92/2.48/910ms.

The resulting scans for 25 patients were then pre-processed using Automated Lesion Identification in Statistical Parametric Mapping software (SPM12: Wellcome Trust Centre for Neuroimaging), running on MATLAB R2017a (The MathWorks, 2017). Images were spatially normalised into standard Montreal Neurological Institute (MNI) space and segmented into tissue classes using unified segmentation with default parameters (Seghier et al., 2008), followed by smoothing with an isotropic kernel of 10 mm at full-width half maximum. Images were then entered into a multiple regression analysis to

correlate pre-therapy brain structure with post-therapy treatment outcomes. The statistical model is reported in Chapter 3.

The pre-processing steps carried out in Chapter 3 were as follows:

(1) Spatial normalisation

Individuals vary with regards to the size and shape of their brain, and their position within the scanner. Before any statistical analyses are carried out at a group level, individual MR images must be standardised to bring them into a common space. Using ALI, the MR images were spatially normalised into Montreal Neurological Institute (MNI) space using the canonical template image in SPM12. The modulation parameter was switched on to preserve information about the amount of tissue present in the original MR image before normalisation. Within this automated process, the individual MR image is normalised to the template in two stages: (1) Affine transformation moves the MR image to match the desired template, whilst preserving the ratios and distances between the parts of the images. It does this by estimating a set of 12 parameters to find the optimal transformation based on a Bayesian framework; (2) Coefficient estimation accounts for global and non-linear shape differences between the MR image and the template, and finds the optimal model which minimises the squared difference between the two images, and maximises the smoothness of the deformations.

(2) Segmentation

Following normalisation, images were segmented into different tissue classes. In ALI, this produced four tissue classes: grey matter, white matter, cerebrospinal fluid, and non-brain. Within ALI, estimation of different tissue classes is based on three main components: (1) A priori probability maps contain information about the spatial distribution of different tissues classes in healthy individuals; (2) Models contain information about the intensity distributions of different tissue classes; (3) Correction for intensity variations which relate to positioning within the head coil in the scanner. This process results in a probability value at each voxel for a particular tissue class, between zero and one, based on the segmentation model.

(3) Smoothing

Following segmentation, grey and white matter images were smoothed using an isotropic Gaussian kernel. Each individual voxel takes on an average grey or white matter value from its neighbouring voxels, based on the size of the kernel, which serves to normalise the data, allowing parametric statistical models to be applied. Smoothing evens out minor anatomical variability which is not resolved during normalisation, and also evens out minor imperfections which occur due to the normalisation process. The width of the smoothing kernel is specified, such as 10 full-width-half-maximum (FWHM), which means 10 pixels in the x and y direction. The resulting image consists of resolution elements (resels) which are expressed as the number of voxels per resel. These resels are treated as independent of each other, and so determine the number of statistical comparisons. In this way, smoothing can reduce the number of comparisons, enabling more lenient corrections for multiple comparisons in the statistical model. The value of the smoothing kernel is chosen based on the size of the regional differences in volume which are expected from the data. In the present chapter a 10mm full width half maximum (FWHM) kernel was applied.

4.2.10.4 Chapter 4: Longitudinal VBM

Serial imaging data benefits from increased power due to its within subject design, and permits demonstration of causality due to intervention, by enabling structural brain changes to be correlated with behavioural measures. Previously, changes in grey matter volume have been observed in response to learning to juggle (Draganski et al., 2004), demonstrating that longitudinal MRI can be sensitive to subtle training related changes in brain volume over time. In the present study, serial longitudinal registration in SPM12 was used to investigate change in brain structure related to Listen-In treatment. This method has previously been employed in studies with neurodegenerative patient populations (Binney et al., 2017; Minkova et al., 2018), and has been shown to be sensitive to atrophy in language regions in patients with primary progressive aphasia (Mandelli et al., 2016; Santos-Santos et al., 2016). In persons with aphasia, one previous study utilised this method to investigate changes in functional MRI in response to language treatment (Martin et al., 2017). To the authors knowledge, the present study is the first study to use

this method to the investigate treatment related changes in brain structure in persons with aphasia.

A major methodological issue which arises when using three serial scans, as in the present study, is additive bias, where volume change can become biased towards the first time interval. Longitudinal registration in SPM reduces this bias in its model, by combining rigid-body registration, diffeomorphic registration, and intensity non-uniformity correction (Ashburner, 2013). In the present study, this method was used to generate divergence maps for treatment and standard care blocks, for each patient. Values within these maps indicate either compression (<1) or expansion (>0) at each voxel. These can be broadly inferred to represent 'loss' or 'gain' in macroscopic morphology of brain matter; however, the underlying neurobiological changes cannot be inferred (Tardif et al., 2016). Divergence maps were combined separately with grey and white matter images to create two probabilistic images for each patient, for grey and white matter. These images were used in later regression analyses to correlate regional changes in brain matter with behavioural treatment effects. The statistical model is reported in Chapter 4.

The following steps were conducted for each patient (see Figure 4-12):

(i) Three unprocessed MRI scans for each time point (T2, T3, T4) were entered into the longitudinal registration toolbox in SPM12 using default parameters. Timepoint for each scan was entered as a decimal of a year which differed according to individual patient scan times. This produced three divergence maps and an average scan image for each patient. The third divergence map was inverted to match the polarity of the first divergence map.

(ii) Divergence maps were subtracted to produce two divergence maps corresponding to volume change over the therapy block and standard care block. For the therapy block, this was T3-T2 and T4-T3 for Group 1 and Group 2 respectively.

(iii) The resulting map was divided by the individual scan interval for each patient, to scale observed differences relative to time across patients (scan interval=days between scans represented as decimal of a year).

(iv) Both divergence maps were multiplied by the patient's grey and white matter images to generate probabilistic 'change' maps for statistical analysis. Grey and white matter images were patients' average scan across three scans. These were generated by segmenting the average scan image (generated in step (i)) using Automatic Lesion Identification in SPM12 with default parameters).

(v) Therapy and standard care change maps were subtracted to produce two single grey and white matter change images for each patient. Each image represented change over therapy more than change over standard care.

(vi) Images were spatially normalised into MNI space using the deformation field from the normalised average scan image.

(vii) Images were smoothed with an isotropic kernel of 6mm at full-width half maximum using SPM.

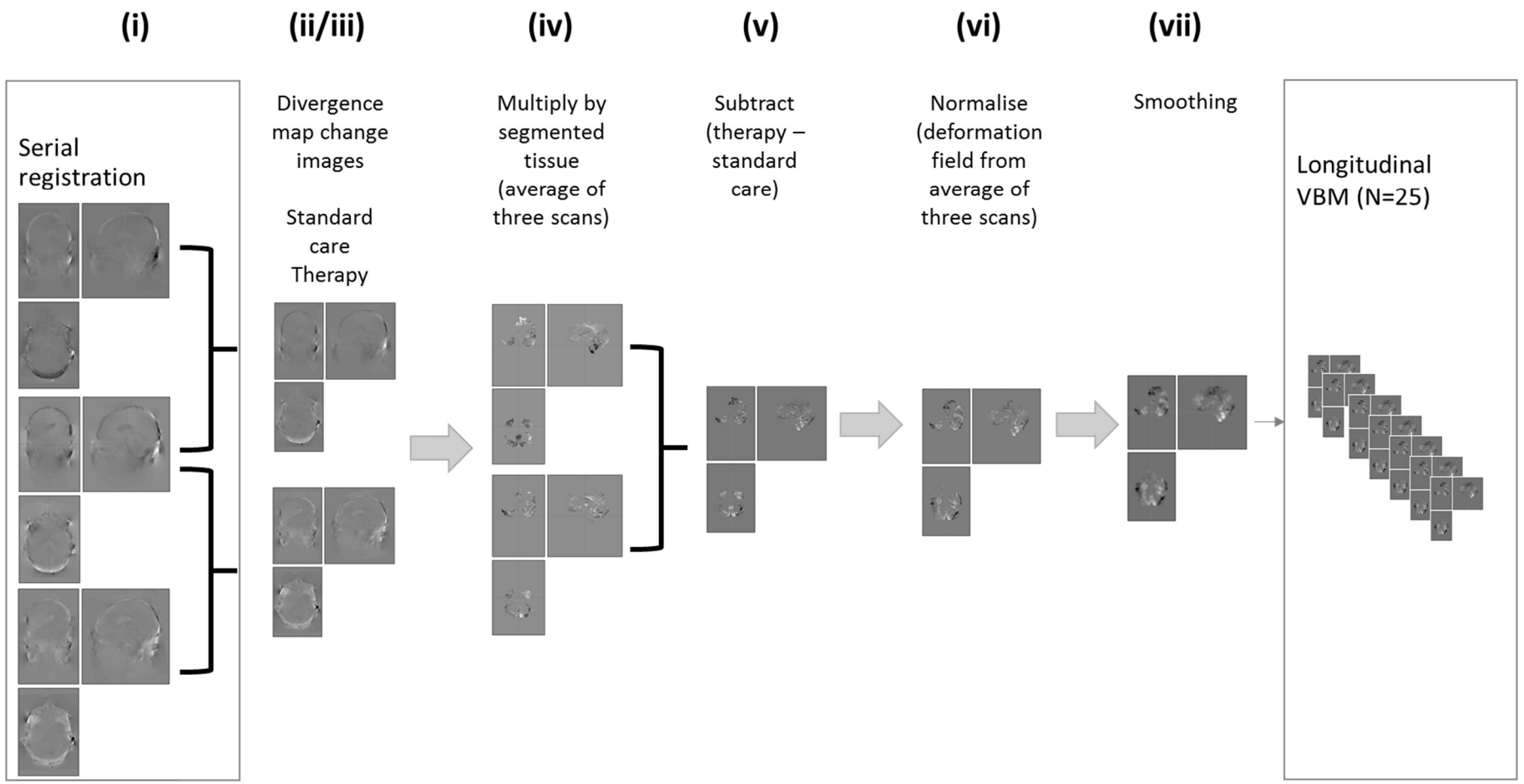


Figure 4-10 MRI processing pipeline for one patient prior to longitudinal voxel-based morphometry.

5 Results

5.1 Chapter 1: Developing a digital speech comprehension therapy app with gamification - A qualitative study

Aim 1: To develop a speech comprehension therapy application ('Listen-In), with gamification, using co-design methods with persons with aphasia.

Objectives:
(1) To carry out five focus groups with individuals with aphasia, to obtain feedback on a series of Listen-In prototypes
(2) To collect qualitative data from five focus groups, and analyse this data to identify key issues surrounding usability and enjoyability
(3) To develop design recommendations for each key issue, and implement these into subsequent iterations

To achieve these aims and objectives, I collaborated with a team of software developers, and persons with aphasia, in an iterative cycle of app design and development. I ran a series of focus groups as a platform to obtain feedback about the app during its development stage, and using qualitative methodology, identified themes in user feedback and translated these into key findings and recommendations, which resulted in tangible design changes for the next iteration of the app. These themes form the basis of the results and discussion sections in the present study.

5.1.1 Qualitative study design

5.1.1.1 Procedure

I obtained signed consent from all participants prior to the first session commencing. Ten participants were able to attend sessions independently; two attended with carers. Each

group was held in the charity's communal space which was familiar to participants, and lasted approximately two hours. Groups were formed of either one group, or two smaller groups, depending on number of participants. This was to facilitate communication by reducing noise, number of people in the conversation, and allowing individuals more time to respond. Connect staff members performed moderator, facilitator and communication support roles, and further communication support was provided in some sessions by SLT students. Games developers were present in some groups to gain insight into our patient population and their specific needs, and to assist in resolving any bugs in the application. The structure of the sessions varied between groups, but all contained the following core components: (1) introduction to the session by moderators and facilitators; (2) a period of playing time in a small group, in pairs, or independently, sat around a table; (3) feedback session following a pre-planned questioning route; (4) period of home play (after groups two and four). The questioning routes covered pre-selected topics targeting key aspects of the app, and were developed in collaboration with the software developers (see Appendix 1). The approach diverged from traditional focus group methodology as supplementary methods were included to support participants communication. These included visual analogue scales, total communication strategies, closed questions, and supported conversation techniques. I considered data saturation and theoretical saturation, and addressed this by using findings from completed groups to develop questioning routes for future groups.

5.1.1.2 Data analysis

I collected data using digital video recordings, visual analogue scales, and notes from moderators and facilitators. The latter two sources were not analysed further as they did not provide additional feedback over and above the video recordings. I followed a thematic content analysis approach (Braun & Clarke, 2006) based on video recording transcriptions using NVivo 10 software. The analysis proceeded using the following steps for each focus group, and formed part of a collaborative piece of work with two further colleagues:

(1) I developed a draft coding framework, consisting of root codes and rules for preselected categories (e.g. 'positive' comments). This emerged initially from focus group one, and was updated and refined iteratively as and when I encountered new areas of feedback in the transcriptions

(2) I transcribed all video recordings, including pertinent non-verbal behaviours. I described these objectively and noted if using any subjective interpretation. I freely interpreted behaviours which showed a clear meaning, and those that were identified as ambiguous were interpreted alongside colleagues.

(3) I assigned code(s) to lines or chunks of transcription based on the coding framework, as a way of organising the transcriptions into categories (e.g. "jigsaw puzzle", "pinball game").

(4) I then ran a coding comparison on a subsection of transcript that was transcribed independently by all team members. Codes which had <80% reliability were considered to show too much variability, and were jointly discussed, to resolve differences. The coding framework was updated accordingly.

(5) Remaining portions of the videos were divided amongst the team and coded independently

(6) As a team we analysed nodes (all text which is assigned the same code, e.g. "sound effects") to extract key findings. From these findings, we made recommendations for changes to the app, which were translated into the next prototype design.

(7) Finally, overarching themes were identified as a team.

5.1.2 Results

Five focus groups were successfully completed, meeting Objective 1. Eight key themes emerged from thematic content analysis of qualitative data acquired across all five focus groups, and the key findings from each theme are described below (Objective 2). Key changes which were then implemented into the app are described, to illustrate how these findings were translated into design decisions (Objective 3).

Listen-In component	Theme	Focus groups
Hardware and software	1 Usability of the hardware	FG1
	2 Audio problems	FG1
Therapy aspects	3 Therapeutic elements	FG1 / FG2
Gaming aspects	4 Understanding gaming elements	FG1-FG5
	5 Gaming enjoyability	FG1-FG5
	6 Metagame	FG1-FG5
Aesthetics	7 Graphics and sound effects	FG4
Other	8 Gaps in user feedback	FG1

Table 5-1 Themes in user feedback derived from thematic content analysis

Theme 1: Usability of the hardware

This theme summarises interactions between users and the hardware, including headphones, charger and the tablet and its touch screen elements (buttons embedded within the touch screen device). In FG1, many found it hard to find the power button, headphone jack, and volume buttons initially. Some needed assistance, but following this, were able to use these independently. Two participants in the group needed help to set up the headphones due to hemiplegia. One participant had difficulty touching specific buttons and pictures on the screen due to difficulty with fine motor control. The touch sensitivity areas for these buttons were made larger in the next iteration in FG2, and the picture size was increased, to allow more room for error when selecting an image. In FG1, many participants experimented with different ways of touching the screen to interact with the app. This included different types of touch, such as swiping and tapping, pressing for long durations, and applying increased pressure. In one observation a participant was uncertain about how to press, and where to press, to interact with the game:

> *Participant: [Observes dropper being filled with coins. Swipes finger on screen several times. Nothing happens. Mouths 'I don't know?'. Swipes finger several more times. Taps once. Coin drops. Turns to moderator, shakes head, holds out hands in exclamation. Continues to touch jigsaw buckets one at a time to try to release coin]*

By the end of FG1 all participants were able to use touchscreen without assistance. At the end of FG2 participants took the tablets home. When they returned in FG3, two participants reported not carrying out any practice due to difficulty waking the tablet from 'sleep' mode, and because the tablet repeatedly crashed. Two participants with hemiplegia reported using the in-built speaker as they were unable to use the headphones independently; during sessions, they used headphones with support. No other difficulties with the hardware were reported. For the second home practice session, troubleshooting sheets were provided.

Theme 2: Audio problems

This was a discrete theme which occurred in FG1, and related to the audio volume of the spoken stimuli being too low. It dominated feedback from participants in FG1. In many cases, participants were unable to respond to the task as they could not hear the audio. Following FG1 the research team identified a more suitable tablet with good quality external speakers, to increase volume, and meet the needs of individuals unable to use headphones independently at home. In FG2, after the volume was rectified, some participants still reported difficulty with the audio. One participant reported that the voice sounded muffled:

> *Participant: "More sounds, more atmosphere, more [unintelligible] it's the voice, um, listening to the voice, um, muffled"*
> *Facilitator: "Did it sound muffled?"*
> *Participant: "Yes, but, the listening, clear, but the voice"*

For one participant, the difficulty appeared to be due to discrimination of two similar sounding words (a minimal pair):

> *Participant: "I pressed [unintelligible] but I wasn't sure if it was saying ball or wall"*

Following this feedback, audio recordings were developed in collaboration with a company specialising in digital audio production, and were recorded in a professional studio, to maximise audio quality of therapy challenges.

Theme 3: Therapeutic elements

This theme incorporated all aspects related to the therapy component of the app (spoken word-to-picture matching). The key findings which emerged concerned the quality of the audio stimuli and pictures, clarity of therapy feedback, and novel therapy elements (e.g. coin rewards).

Quality of the audio stimuli and pictures were generally acceptable to the group. Some participants identified pictures which they thought didn't match the target (*e.g.* "police" → sheriff), resulting in a robust audit of the therapy challenges by the research team, and a patient expert, to ensure pictures were appropriately matched. Some participants commented that they liked the quality of the images used. Many participants found the audio too low in FG1 (see Theme 2), but in future groups, did not raise any further concerns.

The therapy task appeared intuitive to participants, and no difficulties were observed. One participant became agitated and frustrated, which appeared to be due to finding the task difficult, and wanted the written word to be added. In later focus groups, many participants found the therapy task repetitive:

> *Participant: "I've already mentioned it [makes circles with gingers]. Repetition. Constantly repeating the same questions".*
>
> *Participant: "The questions are a bit repetitive"*
> *Participant: "Right right right" [nods head]*

In this prototype there were a small number of therapy questions, and these were increased in the next iteration to provide more variety. (In the trial version, a large number of therapy stimuli were included, so this was not envisaged to be a significant problem).

When asked directly, some participants agreed that the game would encourage engagement with the therapy task.

Theme 4: Understanding gaming elements

This theme arose from observations and feedback relating to difficulties with the individual gaming elements in the app (separate from wider understanding of gamification, see Theme 6 Metagame). This included problems understanding the individual mechanisms (e.g. coin release), as well as problems understanding the purpose of these elements.

Coin collection

Participants collected coins on a trial-by-trial basis within the therapy task. Some participants did not notice this occurring on the left-hand side of the screen. This was an important gaming element which linked the therapy to the game (see Theme 3: Therapeutic Elements), therefore, following this observation, it was re-developed into a transparent 'pot' where participants could see the coins being 'won' trial by trial, as well as the total number of coins they had collected (Figure 5-1). A visual animation was also introduced to draw in participants attention. In later focus groups, no further difficulties with this gaming element were apparent.

In the gaming component, one participant mentioned that the buckets at the bottom screen (which the coins fell into) did not adequately show how many coins had been collected:

> *Participant: "…So you could spend all your coins on one bucket and not know how many it is to fill one?"*
>
> *Participant: "…it got quite annoying when you got the puzzle piece and the coin dropped into the same bucket, so it was wasted"*

Following this, the team implemented a clearer visual effect to indicate when particular buckets were full (Figure 5-2).

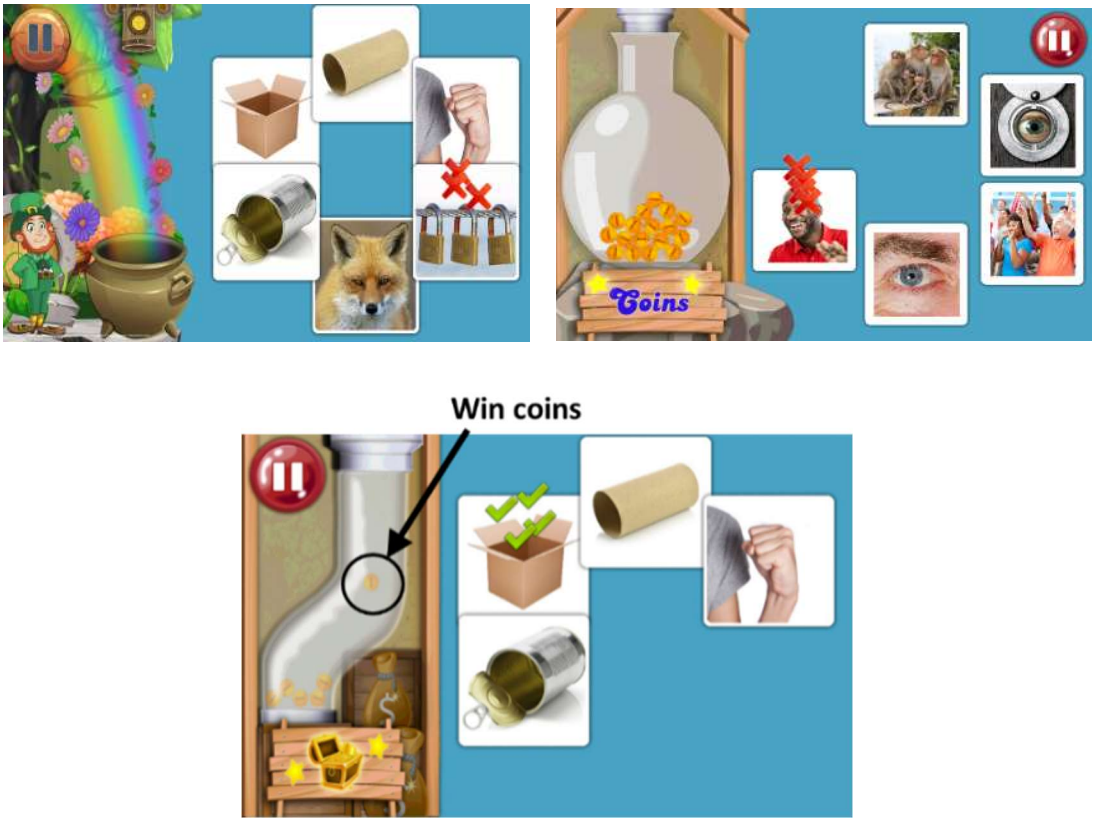


Figure 5-1 Screenshot of an early (top) middle (left) and late (right) stage prototype, demonstrating changes to the coin reward container.

Figure 5-2 App screenshots showing an early (left) and late (middle and right) version of the coin bucket.

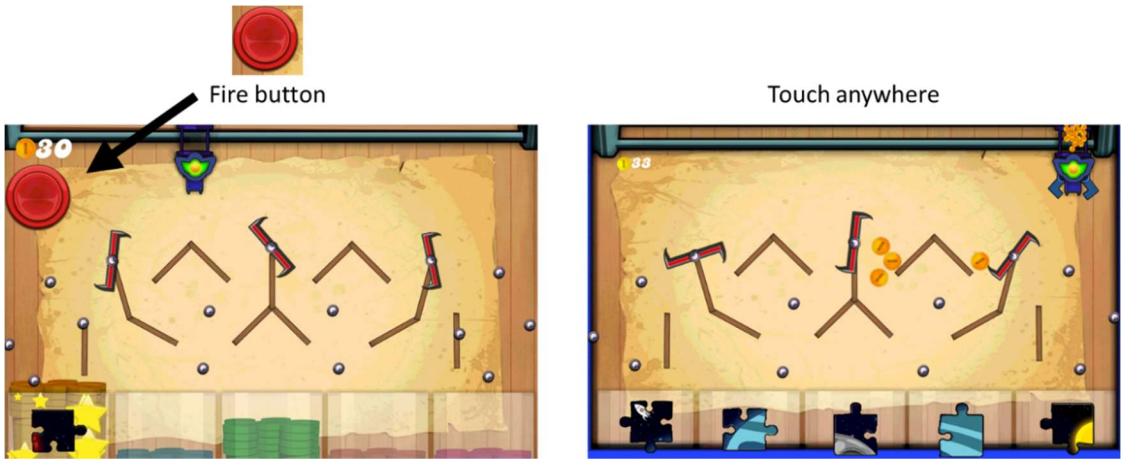


Figure 5-3 Screenshots showing the app with and without a fire button, a discussion point which arose in FG1/FG2.

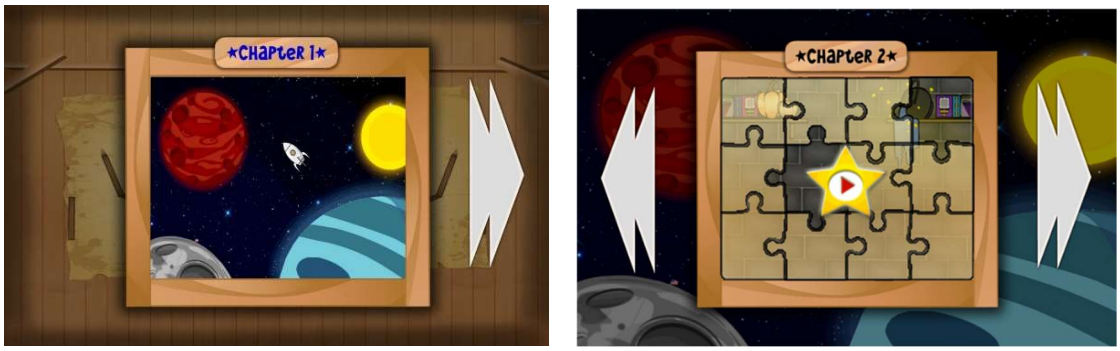


Figure 5-4 Screenshot of an early prototype showing arrows to transition between levels.

Releasing coins

Early difficulties with mechanisms mainly concerned how to release the coins in the game. A further observation was that participants did not, at first, understand the point of the game, where the goal was to release the coins to aim for a specific jigsaw bucket (Figure 5-3). In FG1 some participants were not sure where to press on the screen, and this was rectified by adding a designated 'fire' button (Figure 5-3). In FG2, all participants were observed using this button independently, and liked the addition:

> *Participant: "Don't matter, I prefer fire [button]"*

In FG5, a new participant was able to find the fire button independently, after some self-directed experimentation:

Participant: "It took me a little while to twig it was a fire button, but then it was fine"

In FG2, two participants thought having two fire buttons on each side of the screen would be helpful, but another participant thought that the "exit" button should remain on the right-hand side, instead of a fire button:

Participant: "I am OK with either one but if you've got limited movements it might be better to having one on either side"

The majority of participants liked the single fire button, and the research and software team agreed the exit button was important, therefore no further changes were made. In FG3, one participant reported preferring the designated fire button, rather than pressing anywhere on the screen, due to the possibility of releasing a coin by mistake:

Moderator: 'So would anywhere on the screen be better for you then?'
Participant: "Erm ... In actual fact ... it would be worse cause ... I think with a fire button I only hover near it when I want to use it, I keep it away from it anytime I don't want to use it [another participant agrees] and but, if I, had an involuntary muscle spasm, it would touch it anywhere on the screen, it would drop it anyway. So, but, I can see advantage of that.'

In FG1, some participants felt that it was unclear how far along they were in the game, and one suggestion was to include a counter which indicated the number of coins remaining to be released. This was implemented into the next iteration of the app.

Multiple levels

Only one participant progressed to a new level in FG1, highlighting that the arrow feature (which was implemented to allow navigation between levels) was not obvious for most participants. There was a lack of feedback on this feature in further focus groups.

Aim of the game and use of strategies

In FG3, following sufficient time with the application for familiarity to occur, many participants still did not understand the overall mechanism or purpose of the game. One participant showed considerable frustration, and at home, one participant's wife also didn't understand the task:

> *Participant: "I don't know… [Shaking head] Where? Where? Where? But where? What's that? Where?"*
> *Therapist: "So-"*
> *Participant: "Where?" [looks around screen and appears confused about where the coins come from and how to release them]*
>
> *Facilitator: "So it would be nice to have some support?"*
> *Participant: "I got my wife to look at it and she couldn't do it either. Where's the instructions!?"*

A connected finding was that some participants still did not appear to use strategies when timing the release of their coins, to maximally win jigsaw pieces. This was demonstrated by many participants releasing coins in their own time, and not in response to obstacles in the game. One participant did appear to consider timing when releasing the coins, and also showed a clear understanding of the game, highlighting a connection between use of strategies and wider understanding of the goal:

> *Screen: [Game finishes, jigsaw screen appears, no jigsaw piece has been won]*
> *Participant: "Obviously we... [navigates between frames using white arrows] obviously we didn't do enough [smiling] because we've got to repeat the questions" [presses play button]*

In FG4 and FG5, the same pattern of variability occurred amongst participants. Some appeared to engage with timings so they released their coins to either to directly fall in a bucket, or avoid or hit a particular gaming element; whereas some participants pressed the button quickly, to release all the coins as soon as possible; and some participants used

a mixture of approaches. For example, in the Alpine Ski level, a participant waited until the skier had passed until releasing the coin.

Game variety and degree of perceived difficulty

Participants found some levels more difficult than others:

> *Participant: 'It's … I like that but it's difficult. I try … yeah yeah …'*

For example, in 'Tricky Lights', the last level, no participant understood the strategy required to win the jigsaw pieces. One participant tried to avoid the crystals, rather than aim for them. In the 'Boating' level, only some participants found out how to open the barricade by hitting a red 'life-ring' target. Several participants found the movement of some of the gaming elements too fast in one level, and this was slowed down in the subsequent prototype:

> *Participant: "Yeah, yeah, it's difficult"*
>
> *Therapist: "Yeah, found that hard [patients agree with murmurs and nods]"*
>
> *Therapist2: "Do you think that the dolls were too fast?"*
>
> *Participants: "Yes" [two patients nod firmly]*
>
> *Therapist2: "Yeah? Too fast? OK"*
>
> *Participant: "Yeah, I think so…"*
>
> *Therapist2: "Too fast?"*
>
> *Participant 1: "…because of the [points to dolls on screenshot of level]*
>
> *Therapist2: "Yeah? Do you think there were too many dolls? Too many of them?*
>
> *Participant 2: "Yes… it's maybe slightly…"*
>
> *Participant3: "I wouldn't say no, I'd just say they're too…"*
>
> *Therapist2: "Too fast?"*
>
> *Participant3: "Yeah"*

Therapist2: "But not too many?"

Participant3: "Yeah, not too many"

Another participant with a good understanding of the app in general had difficulty understanding one particular level. As a result, the design of this level was also changed.

Therapist: "So did you notice that the duck did turn?"

Participants: "Yeah. No [laugh]"

Therapist: "Did you notice that?"

Participant: "But I was trying to avoid the ducks! [laugh]"

These differing levels of understanding and variation in strategies used appeared to be related to individual differences and preferences, rather than a problem with the gaming mechanism itself. In addition, lack of understanding of some aspects did not appear to impede enjoyment or progress in the app as participants were still observed interacting and progressing with ease, and therefore it was not considered a problem. Following this observation, the range of levels were increased in two ways: firstly, by increasing the variety of themed levels, and secondly, by incorporating bronze, silver, and gold levels for each theme, which increased in perceived difficulty, by incorporating additional gaming mechanisms (e.g. more obstacles) (Figure 5-5). This resulted in 63 different gaming levels.

In FG5, participants continued to vary in the nature of their interactions with the game. It was clear from observations at this stage that all participants understood how to navigate the basic gaming mechanism of releasing coins to fill buckets.

Demo feature

In response to some participant did not understanding the purpose of the game, a Demo feature was developed as a 'walkthrough' to guide players through one cycle of the app. In FG5, one participant was new to the group, and quickly understood the gaming mechanism following the demo feature:

Participant: "Yeah – it made me understand a little bit more. What to do now. Yeah 'cause I didn't know what to do with the end parts. You get a jigsaw piece … I didn't know how to get it"
Moderator: "Oh, how to get it. But now, after the demonstration, you understand the steps?"
Participant: "Yeah, I am put it in like, aim!"

Participants already familiar with the application found this part frustrating due to the slow pace of following steps they were already familiar with, but agreed it would be useful for new users:

Therapist: ''So this is just the demonstration''
Participant: [audible sigh]
Therapist: ''Too easy?''
Participant: ''No!''
Participant: [touches the images on the demo version]
Participant: [opens hands outwards in confusion and presses the pictures on the screen]

As a result, the demo was considered important for new patients, but needed to be non-intrusive after initial use. It was therefore included as an optional button on the main home screen (World Map) in the final version.

Theme 5: Gaming enjoyability

Gaming enjoyability included all feedback and observations relating participants reactions to the app. A large proportion of this theme involved participants' confusion surrounding the purpose of the game, which later developed into its own separate theme (4 Understanding gaming elements).

In FG1, 3-4 participants shared a computer tablet, creating a shared user experience. Participants interacted with each other whilst taking turns to play, and this appeared to create a sense of social connection and competition. Due to resource constraints, as well

as for the purpose of keeping the clinical trial consistent across participants, this avenue was not explored further.

In FG1 and FG2, the gaming element consisted of an aim-and-release coin game, with a single theme. Participants appeared to enjoy this aspect:

> *Participant: "I am not a person who will go and play computer games 24 hours a day, but I can see myself playing this game every day, may not be for hours, but I can do it!"*

However, some expressed concerns it was "boring" as there was no challenge to it. Some participants felt the game could be a hurdle to future players being able to complete a high dose of therapy:

> *Participant: "So as long as it's shorter [the game] … 'cause you can't progress to the next level until you…*
> *Therapist: "…until you've done it."*

Two participants appeared to be frustrated by the game, and expressed that they did not understand why they had to release a coin, and its overall purpose. This seemed to negatively impact their enjoyment:

> *Participant: "But because we didn't… we couldn't understand it. We spent a lot of time trying to find instructions. What to do … I still don't know what's the aim of this is.*
>
> *Therapist: "The game is really hard. … Why the game?"*
> *Participant: "Why… bang! Oh! [pretending to play] Why?!*
>
> *Participant: "I can't, I couldn't… I stuck on the game. There is no explanation. And no need to do it"*

In FG3, following a period of home play, most participants appeared to have enjoyed the game, although some still thought it was boring. In response to this theme, the team

developed further levels, and presented these to the group. These received positive reactions from many participants:

> *Participant: "It's good! [tapping fingers and showing the watch] Very good!*

During observations participants appeared to enjoy the new levels and felt they would be appealing to different people. One participant enjoyed the 'hard' levels, but other participants became frustrated with these at times:

> *Participant: "'Cause it [coins] bounces all over the place. It will take me forever. And you get frustrated."*
> *Facilitator: "You were thinking to give up?"*
> *Participant: "Yeah"*

To capture the relative interest for each level, participants were asked to give each level a rating out of 10. A range of ratings were given, and these tended to fall between 5 and 10, indicating an overall positive response; however, these were given within the group so may have been subject to bias depending on other group members responses. A pattern appeared in ratings, as some levels tended to be rated higher or lower than others, indicating that the different levels evoked different reactions.

As a result of this feedback, the team spent considerable time discussing the key features of each level, to ensure there was sufficient variability to retain participant interest over time. The feedback from participants was varied, and showed individual preferences. As a result, further developments focused on increasing the range of difficulty for each level to accommodate a range of player preferences. One suggestion was to increase the perception of task difficulty, rather than increase level of skill required. In FG4, more new levels were presented. Ratings again showed variability in preferences for different levels.

Theme 6: Metagame

The metagame referred to the wider purpose of the gaming elements, and overlapped with the 'gaps in user knowledge' theme, as there was insufficient feedback about participant understanding of the wider purpose of the game, and how each part of the

app was interconnected, particularly in the earlier FGs. In particular, there was insufficient feedback about the jigsaw puzzle. This was an important part of the app, as it has been developed to fulfil a 'motivational' role to keep patients playing to 'collect' pieces, and thereby increasing engagement and dose.

Participant: "I don't get what the object of the game is"

As a result of this consistent finding, one recommendation was to add transitions to each screen as a way of linking the different parts together, and creating concrete links between each screen. For example, at the end of a therapy block after users had collected their coins in the receptacle, an animation was added to show these moving into the coin dropper, when then formed part of the next gaming screen (Figure 5-6).

By FG4, although participants were able to interact competently with the game mechanics and were familiar with the repetitive cycle of gameplay, it was still not clear whether many understood the point of the game. As a result, a 'World Map' screen was developed between FG4 and FG5 (Figure 5-7). The purpose of this addition was to act as an overarching structure which provided clarity and direction on the overall purpose of the game. In this way, it was hoped that participants would better understand the idea of playing therapy to travel around the World Map to 'collect' a series of jigsaw puzzles. The 'collecting' principle was the key motivator in gameplay, and it therefore was hoped the World Map would provide clarity on this aspect.

As this feature was added late in development, participants were only presented with this addition in FG5, therefore feedback was based on short periods of play. In general, participants were able to navigate the screen successfully after some familiarisation. The application allowed participants to choose any location on the World Map in any order, and no problems were observed interacting with this set up. However, some feedback indicated that participants needed clearer direction on where they currently were on the map (represented by a Leprechaun character with an arrow), and where they were supposed to be going next:

Participant: "It's OK. But it should show you which direction…"

Moderator: " … you prefer to be guided? Because at the moment you can do anything you want to"

Participant: "Yes, you don't know what's going on"

Participant: "I don't know why he's going that way [leprechaun arrow]"

As a result, the team implemented a designated route around the map in the final version, indicated by a red arrow, to provide clarity on location and direction. A further piece of feedback was that the World Map screen appeared too often. In this prototype version, patients only had five questions per therapy block to maximise time spent interacting with the World Map. The final version had 15 question per block which therefore increased this time window. In summary, the metagame appeared to be well understood by all participants after a short period of home play. Some minor adjustments needed to be made to increase clarity on navigating the World Map.

Figure 5-5 App screenshot demonstrating bronze, silver and gold difficulty levels for one level.

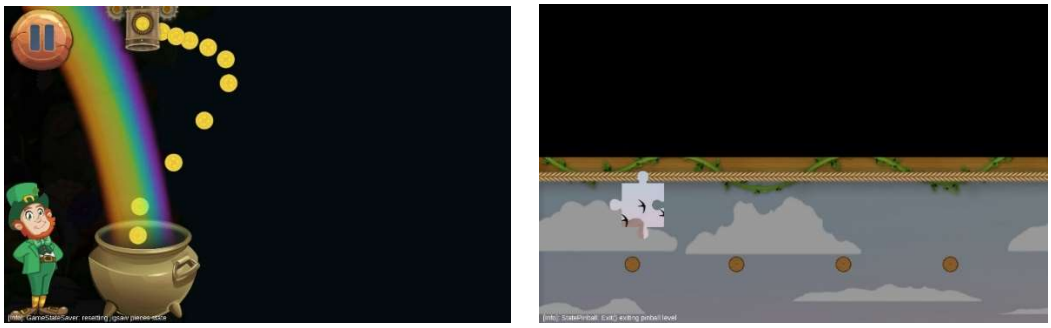


Figure 5-6 Screenshots showing: Left: 'flying' into the coin dropper. Right: Jigsaw 'flying' towards the next screen

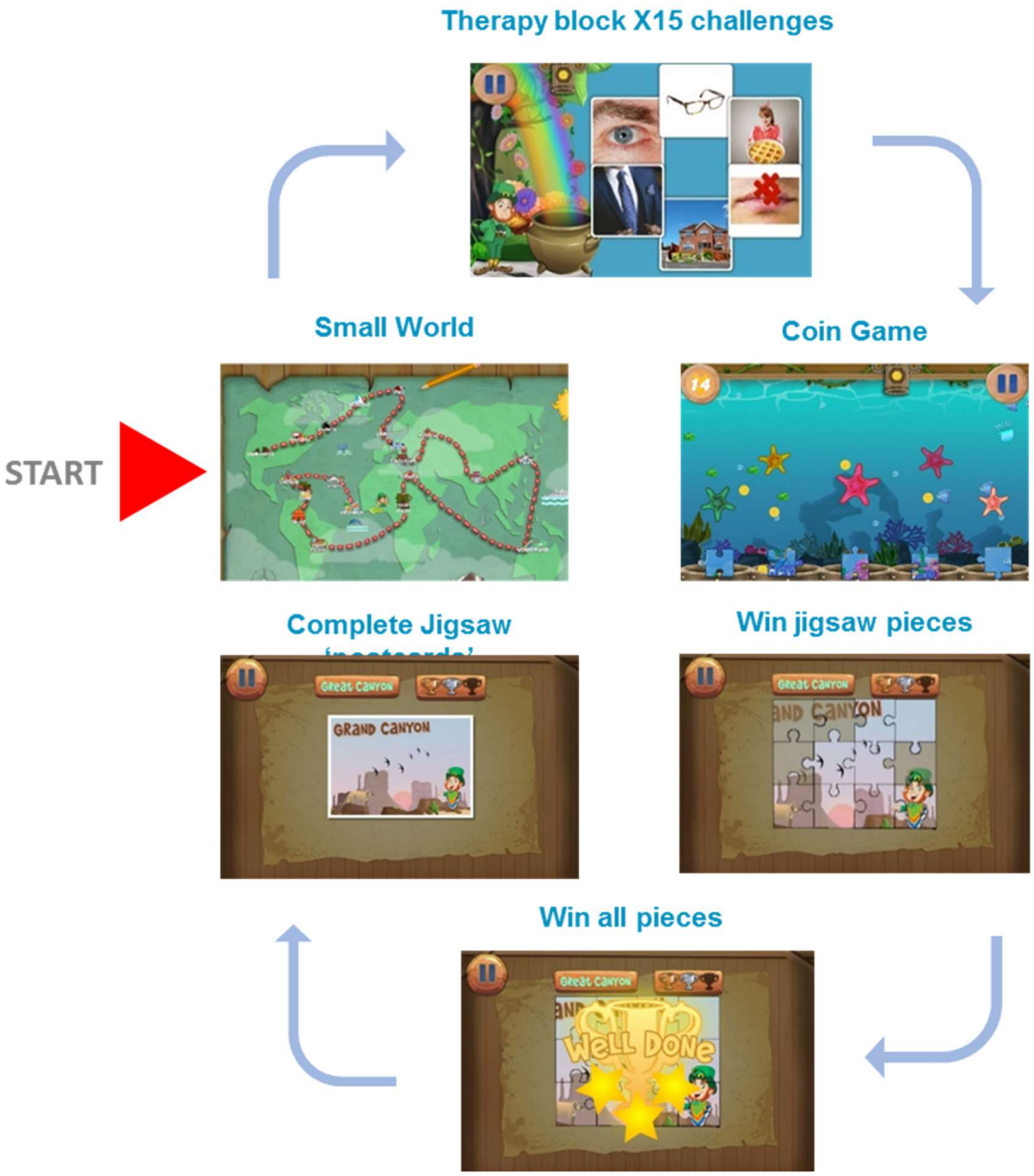


Figure 5-7 Diagram showing progression through each part of the app in the final version of Listen-In

Theme 7: Graphics and sound effects

This theme summarises the results regarding the graphics and the quality of the sound. Considerable time and expense was used in collaboration with the software team and audio company to develop graphics and sound effects which were looked professional, and were engaging and interesting for participants. Findings mainly emerged from FG4 where participants were directly asked for feedback in the questioning route. In general, positive comments and reactions regarding the sound effects were given:

Participant: "the sounds are very catchy"

Participant: "Oh yeah, I like this music [laughs]" … "I like the music on the monkey, because he's like drum [motions banging drums]"

One participant (with a background in graphic design) reported that the backdrop for two games ('Cookie monster' and 'Dodge ball') was "too beige" and therefore not very appealing. One participant pointed out that it was difficult to see the number of coins required to win a jigsaw piece, as the graphics for the coins buckets were unclear. As a result, the bucket graphics were adapted with further animations which indicted the number of coins that had been collected, and whether or not a bucket was 'full'. Another finding that emerged in this Focus Group was that the graphics could be potentially distracting, or cause confusion in the game. For example, on the 'Halloween' level it was observed that too many 'ghosts' could be distracting, and on the 'Funfair' level the moving element did not aid understanding about the strategy required. The team therefore decided to manipulate the number of moving graphics components to make the games easier/harder for participants as appropriate. In general, there was limited direct feedback concerning the graphics and sound effects, perhaps because participants were generally familiar with these by FG4. Many participants responded positively to different levels, and this was likely to have been significantly impacted by the quality of the graphics.

Theme 8: Gaps in user feedback

During transcription and analysis, gaps in feedback emerged relating to individual participants in the group, or certain elements of the application. These were assigned into a unique theme, as they formed an integral part of the research process, resulting in direction and focus for next group. FG1 transcriptions were weighted more heavily towards those with greater expressive language skills. Although efforts were made to include non-verbal transcriptions, these were inherently less interpretable. Therefore, a key change in subsequent focus groups was fully utilising a range of total communication strategies, to enable feedback from all participants to be collected. For example, use of rating scales and visual aids to support and capture feedback from all participants.

Emphasis was also placed on asking every participant for feedback, and therefore the focus groups became increasingly structured. For example, facilitators were asked to direct open questions to every member of the group to ensure all participants had an opportunity to respond.

In FG1, feedback on certain aspects of the application had not been captured. This included the relative durations of different parts of the app (*e.g.* therapy task and mini game), and understanding of the overarching gaming element. For example, one participant touched on this point: “I don’t get what the object of the game is” but no further feedback was identified. Due to the limited time constraints within the focus group, two home play periods were introduced to the development stage to allow participants to use the app for extended periods of time, in their own home. This increased the quality of the feedback, as initial difficulties (e.g. where to press) were overcome so that feedback could focus on deeper issues (e.g. understanding of the overarching game, as described in Theme 6: Metagame).

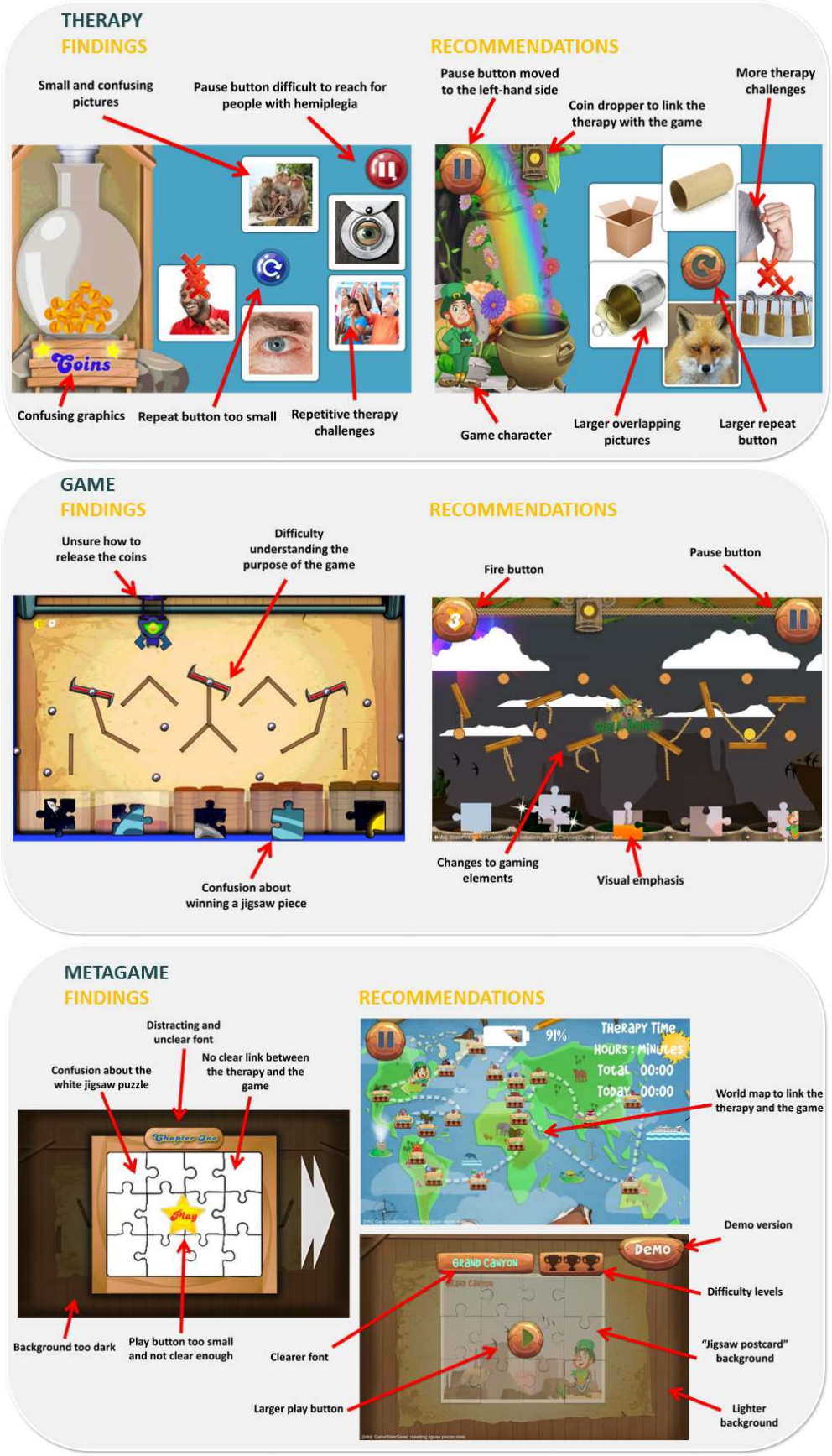


Figure 5-8 Summary of key findings and recommendations from focus groups

A selection of quotes which provide insight into patient's view on taking part in focus groups:

'it's hard, (expletive), but so important'

'Isn't it obvious? You have to involve us in the design phase, because we will be the ones who end up using it'

'If you don't involve us in the development then how do you know if it's fit for purpose?'

'to start with it was no good, then gradually, gradually, to the end it's better. It was useless in the beginning but then we made it better, you need us to make it better'

'sadly, I have lots of spare time and it's fun and interesting'

'it's so important, everything so far is Ok but it's boring, you need people who know how boring it is or it won't work'

5.1.3 Discussion

Eight key themes emerged from qualitative analysis of the focus group data. The content of these themes evolved as the iterative cycle of development progressed, and as different individual elements were considered by the group. The key findings in each theme were used to guide the questioning routes of subsequent focus groups, and enabled a large number of design changes to be implemented into the application which were based on direct feedback from individuals with aphasia. These findings highlight the importance of involving patients in the design of digital applications in early stages, so that key decisions are not based solely on the views of researchers or software developers. This was found to be particularly relevant given the novelty of digital gaming in this particular group of aphasic individuals, and is supported in research addressing gaming in older adult populations: "active engagement of older adults in the design process is imperative to successful take-up of the technologies, bridging the generation-gap of young creative and older users'' (van Bronswijk, cited in Marston, 2013). In the case of Listen-In, this was clearly exemplified by the pervasive and unexpected finding that many participants did not understand the overall purpose of the game.

Useability

An essential component of ascertaining initial usability was 'free play' at the beginning of FG1. This provided valuable insight into the likely success of future patients being able to successfully, and independently, use the application for the first time. Many participants needed support with both hardware and software components at first, such as finding the volume buttons, and where to press on the screen; and two participants reported some hardware issues at home (e.g. waking from sleep mode). As many had not used mobile or tablet devices before, implementing home play enabled participants to become familiar with the device, and the app, in their own time. It was clear as focus groups progressed that all participants become more competent with the tablet and navigating the application's interface. Overall, participant's mastery occurred quickly and with apparent ease following minimal initial instruction and a relatively short period of home play. The

qualitative data support this observation, as transcriptions from later focus groups centred around conceptual themes rather than hardware or software related issues. Enabling all participants to master the app was an important goal in terms of Listen-In's usability, as low self-efficacy with technology can be a barrier to sustained use in less experienced users (Marquié et al., 2002), and could be particularly challenging for individuals with language impairments. Other studies investigating tablet-based app use with individuals with aphasia have reported similar findings relating to the need for hardware and software support with this population (Kurland et al., 2014; Mallet et al., 2019). Importantly, in the current study, no persistent difficulties navigating the Listen-In app itself were observed amongst participants. These findings suggest some individuals with aphasia may need support when setting up and using Listen-In for the first time, particularly if using a new device. However, the high degree of competence of all group members, who had a range of aphasia profiles and technology experience, suggests Listen-In is intuitive and easy to use. 'Audio problems' were a major theme of focus group one, but were quickly resolved. This highlighted the careful consideration required when choosing consumer technology for populations with additional requirements; in this case, a tablet was identified which had high quality external speakers.

Graphics and sound effects

The graphics and sound effects appeared to be a major contributor to the overall feel and 'finish' of the game. Many participants appeared to respond favourably to the variety of animations and sound effects. Although it was evident from first-hand experience during focus groups that participants enjoyed the variety of graphics and sound effects, this was somewhat lost during qualitative data analysis, as there was limited verbal and non-verbal feedback directed at these specific components. One solution in future groups would be to record first hand impressions as an onlooker using field notes, which may provide some alternative evidence for findings which are less tangible in video data and transcription.

Gaming mechanisms

Three themes emerged which related to the gaming aspect of the application, and concerned the individual gaming elements ('Understanding gaming elements'); the overarching game ('Metagame'); and how enjoyable participants found the game ('Gaming enjoyability'). Again, participants quickly learnt how to interact with the game after just a short period of familiarisation in the focus group, followed by extended play at home, which again appeared to reflect gradual mastery of the application. Participants varied in how they interacted with each game; some used strategies, indicating an understanding of the end goal to win particular pieces, whilst others played without strategies. These disparate styles and levels of interaction posed a particular challenge in how to develop the gaming mechanism: in order to remain accessible to a range of individuals, the game needed to have minimal cognitive and memory requirements, whilst being sufficiently appealing for those who wished to engage more actively with strategies. In response to this, a variety of themes were developed, which had a common underlying gaming mechanism, each with three level of difficulty. These themes generated different types of reactions: some laughed in response to particular gaming elements (e.g. cookie monster), or particularly enjoyed certain sound effects, and some levels were reported to be easier than others. These reactions appeared to reflect individual preferences, not unexpected given the heterogeneous nature of the group.

As development progressed, the focus around gamification shifted from understanding of discrete components (e.g. coin reward), to how these components were interrelated to form a game with a clear overarching goal. It became evident that although the links between the therapy and game elements were clear to the research team and software developers (and even considered 'simple'), many participants did not understand the purpose of the game. There appeared to be a number of reasons why individuals had difficulty understanding these links. All participants were significantly less familiar with the app than the research team. For some, extended play at home solved this issue; however, some participants still failed to see the point in the game. Some seemed to be generally unfamiliar with gaming concepts due to minimal previous gaming experience, and did not appreciate the significance or purpose of 'winning' virtual jigsaw pieces. This

finding was also echoed by a participant in a focus group study of older games: “I play a game, but I do not understand it. I win points, but totally do not understand how…” (Nap et al., 2009).

A ‘Demo’ feature was introduced in FG3 with the aim of increasing links between the different screens within the app, and so improving participants understanding of gaming elements. Somewhat surprisingly, many participants became frustrated by this feature as by the time it was introduced they were already familiar with the gameplay; however, they agreed it would be useful for new users. Importantly, for one new group member, the demo clearly served to explain the aim of the game. Previous evidence has suggested that older individuals struggle with fast game speeds (Nap et al., 2009); building on this, these finding suggests that elements which are too slow or simplistic can equally cause frustration to users and become a potential barrier to enjoyment, therefore finding a middle ground, appropriate to the target population, is vital. In Listen-In, this was easily achieved by making the demo component optional for experienced users.

Persistent feedback concerning the point of the game culminated in a key turning point during development, and in implementation of a ‘World Map’, which attempted to bridge together the therapy-game cycle. Feedback was only possible in the final focus group; however, initial reactions appeared to be positive, and observations showed participant navigating it with ease.

Overall, these reactions to the gaming component were a surprise to the research and software team, as it was assumed that all individuals would understand the game, and find it intrinsically enjoyable and rewarding, albeit to greater and lesser extents. These findings suggest that, for some, gaming elements may actually be de-motivating. For example, one participant learnt to re-start the app to avoid playing the game entirely and focus only on the therapy. This highlights the importance of involving patients in the design phase, and the unconscious bias held by researchers and developers which may lead to design decisions which don’t reflect the needs of end users. As the purpose of development was to create a gamified therapy app, and given the majority of participants enjoyed this aspect, the game was not significantly altered. However, considerable time

was spent by the software team designing and developing further levels, to increase variety, and try to meet the preferences of a wider number of potential future users.

Therapeutic task

There was significantly less feedback concerning the therapeutic task (spoken word-to-picture matching), which appeared to be due to lack of any particular issues. All participants appeared to enjoy cycling through the therapy challenges, and the participant who actively disliked the game, was extremely positive about the therapy component. Participants commented on the quality of the images; although some pointed out inaccuracies, such as a sheriff instead of a policeman. This feedback was important, as issues seen as 'minor' by the research team may not be seen as such by future users, and could be a significant barrier to engagement for some patients. A key finding was that members commented on the repetitive nature of therapy questions, in line with previous findings (Kurland et al., 2014; Varley et al., 2016). In the present study, a large number of stimuli were developed (>3000 unique challenges) to try to reduce the impact of the repetitive nature of challenges.

Gaps in user feedback

One gap which became evident from the first focus group was the reduced volume of feedback from members with more severe spoken output impairments (see next section). This was problematic for two reasons; firstly, it risked biasing findings towards those with greater expressive language abilities; and secondly, it risked confirmation bias by the researchers when interpreting non-verbal behaviours from members with less spoken output (I.E. interpreting non-verbal behaviours in line with the researchers own view). In order to obtain feedback which was balanced amongst all members, specific strategies and actions were taken in the later four focus groups to facilitate feedback from all members, such as increased focus on use of visual aids which avoided reliance on purely verbal exchanges.

Use of focus group methodology with persons with aphasia

At their core, focus groups are built on shared communication amongst members, and it is this aspect which separates them from other qualitative methods, such as interviews. However, individuals with aphasia have impairments in this key requirement, therefore a number of observations and reflections developed from attempting to use this traditional methodology with individuals with language impairments. These groups necessitated substantial logistical planning, both in terms of recruitment to the group and planning each session. Running a group with such a range of individuals with different severities of aphasia was a challenge, one which has been previously encountered (Wilson et al., 2015). However, the use of strategies reported by Wilson and colleagues (2015) helped significantly in this respect. Overall, the emphasis for running successful focus groups with PWA seemed to be ensuring there was sufficient flexibility in the design to deviate from standard focus group methodology when needed, either as a pre-planned conscious decision, or in real-time. This was vital in order to facilitate the variety of disabilities experienced by the group's participants. To this end several barriers and facilitators to running focus groups with this population were identified.

Several small usability factors were identified as potential barriers for some individuals, which allowed for clear recommendations. For example, one person with hemiplegia had difficulty with the fine motor movements required to touch the repeat button. Whilst this issue was easily resolved by enlarging the button and increasing the 'buttons' touch sensitivity area easily resolved, considerable time was required, in tandem with more facilitator support being available, to ensure that small hurdles like this were quickly remedied and did not become a barrier to inclusion within the group.

An additional hurdle was quieter participants. These individuals were not reluctant to participate, but were often just unable to verbalise or communicate their contributions in an effective or timely manner, due to their aphasia. This became a particular issue in the data analysis stage where it very quickly became clear that non-verbal descriptions within a transcription were just as important as the verbal ones. That is, the adamant shake of a head from a participant with significant speech production difficulties was as salient as

the participant who offered a short monologue as to why they didn't like a particular element. Again, this often relied on skilled facilitators to identify these subtle contributions and then facilitate them fully. Other essential elements included reducing the number of open-ended questions, specifically to support communication skills, and when using open-ended questioning ensuring that adequate scaffolding techniques were available to support participants. A plethora of non-verbal communicative support was developed and utilised throughout the sessions and a reasonable amount of structure was also included in the focus groups specifically to provide participants with a written format to follow in order to scaffold their communication when appropriate.

The participants offered a wealth of experience and insights which drove and improved the development of the application, but also ensured that the application will be better received by patients in the future. Future research involving PWA can no longer shy away from including PWA directly in research because of their difficulties communicating, but instead demands that they are not just included, but are able to significantly drive the research that impacts on their well-being and rehabilitation. The fact that the majority of video games are played by younger populations (Pratchett, 2005), coupled with the reality that stroke is most common in the over 55 age group, demonstrates in statistical terms why it is essential that gamification strategies implemented within any therapy are specifically developed and tested by their target populations. Feedback from participants was very positive, and many commented on the importance of involving them in the design process. The participants took part in a voluntary capacity in the present study; however, Wilson and colleagues (2015) participants were employed as expert consultants. Future studies may look at including these costs during the grant proposal stage, to ensure that PWA are involved in a significant capacity, and receive remuneration for their expert knowledge.

Software developers were a third key collaborator in the development stage, and were new to the field of aphasia, and indeed healthcare applications in general. Perhaps one of the most important steps in this process was when the gaming team met people with aphasia for the first time and observed some of the difficulties that had been impressed on them by the research team. Until this time the descriptions of the difficulties

experienced had been purely academic, despite the passion of the clinicians. When the gamers observed how challenging and invasive these difficulties could be to communication, they began to not only observe with 'broader' eyes but also be more creative with options that they had previously not felt necessary (such as finding alternative options not based on written language). In this respect, the focus groups served to align the differing perspectives of the research team and developers towards a common goal, which had the end user in sight. Involvement of developers with end users should be a critical component of all application development projects involving specific users.

Limitations

A number of co-design recommendations were incorporated into these focus groups, such as visual aids to support concepts in the discussions; however, these could have been extended further. The prototype was used as they key activity throughout all session, but this restricted feedback to aspects of the pre-existing prototypes. Previously, Wilson and colleagues (2015) report using an activity where cut-out figures were given to participants so that avatars could be created. In Listen-In, a similar activity could have been given to allow participants more scope for creativity. One of the barriers to using this approach was the time and resources this would have taken. In future, technology options could be explored see if this reduces the burden; for example, there are a plethora of free design apps available which may enable quick access to creative resources on a tablet.

broad approach to transcription was used, transcribing all verbal responses and pertinent non-verbal behaviours, but a consistent framework for non-verbal behaviours was not used. It is possible that some non-verbal behaviours were missed, or misinterpreted. A solution to this would be to use a finer grain of transcription with consistent markers for non-verbal behaviours to reduce possible bias and help in providing context (e.g. deidetic) however the quantity of data meant this was not feasible on this occasion.

In typical qualitative analysis, it is possible to re-read transcripts to regain a perspective of the whole. However, due to participants' communication difficulties, the transcripts were often sparse and contained many false starts, re-phrasing, and often jumped around between topics. It was difficult to re-gain an overview when re-reading particular themes, and revisiting whole videos was very time intensive. Therefore, one limitation is that some parts may have been interpreted in isolation, when the bigger picture may have provided a fuller interpretation. Although this challenge is inherent in all qualitative data, it is particularly evident in focus groups with PWA. One suggestion for future researchers is to produce short descriptions of chunks of time (e.g. recap 15-minute chunks), to help in navigating the data, and revisiting certain parts.

The analysis approach focused on content analysis to answer the main research question, which was to investigate useability and enjoyability of iterations of the Listen-In app. Further analysis could investigate group dynamics, by taking a discourse analysis approach, to identify subtle dynamics in a focus group setting. This may provide deeper findings and recommendations which could aid in developing recommendations for this population in the future. For example, the role of carers and friends in these sessions could be evaluated. Most of our participants knew each other and whilst this helped to create a relaxed environment, it may have affected participants' responses (e.g. agreeing with a friend, rather than disagreeing).

Previously, the repetitive nature of these types of tasks have been reported as a barrier to extended use. It would have been beneficial to have included further periods of more extensive home play to investigate the impact of this on enjoyment. Some participants did comment on the repetitive nature of Listen-In, however, there were limited therapy questions in the prototype version due to time constraints. The final prototype was developed with over 3000 different challenges, therefore this may ameliorate the repetitive nature to some extent.

Future directions

One aspect which surfaced early on, but was not explored further, was social interaction. Participants appeared to enjoy the shared experience of game play and the element of competition. This is a potential avenue to build upon which may improve user experience and maintain engagement, particularly given the high dose recommended for impairment based treatments. For example, one possibility would be to include player profiles within a virtual community, so that players could compete in terms of dose, or rewards, and interact by 'gifting'. This social theme has been previously explored in EVA Park, a social environment for persons with aphasia, and was well received (Marshall et al., 2016). Some previous research suggests older adults may prefer solitary games. In one study of gamers between 45 and 85 years, 85% reported being solitary players, and rated social interaction as a low motivator; however, social interaction was also the strongest predictor of playing time (De Schutter, 2011). This may be because most games are targeted at younger adults. Further co-design opportunities which engage persons with aphasia in the design of impairment based apps are clearly warranted.

A further direction may be to explore other types of games which suit individual preferences. For example, previous research suggests that older adults prefer challenge type games, which include puzzle games and digital adaptations of traditional games (De Schutter, 2011; Pratchett, 2005). In the present study, not all participants enjoyed the coin game; the option to choose a different game could provide scope to meet individual preferences, particularly in light of the heterogeneous nature of the aphasia population. Given the rapid developments of digital technology over previous decades, a particular consideration when looking forward is how gamified therapies may be adapted and developed to suit the needs of different generations with different experiences of digital devices and games.

Final comment

A number of potential barriers to Listen-In were identified and remediated based on direct observation and feedback from end users. The heterogeneous nature of a typical aphasia population, as demonstrated here, highlights the importance of considering a range of individuals when designing digital applications. This may be achieved, as in the present study, through involving participants in the design process from early stages. Collaboration with individuals with expertise in the complexities of human-computer interaction is also warranted, so that scientific principles of therapy may be integrated into an application which also adheres to key principles of good software design and gamification. The current study has only begun to explore the range of possibilities which may be explored in future digital applications.

5.2 Chapter 2: Investigating response to Listen-In treatment in patients with chronic aphasia

Aim 1: To investigate whether Listen-In can improve comprehension of spoken words in patients with chronic aphasia.

Aim 2: To investigate individual treatment effects.

Aim 3: To explore whether baseline factors can explain treatment outcomes.

Hypotheses:

(1) At the group level, patients will significantly improve in their comprehension of spoken words trained during Listen-in treatment.

(2) At the individual level, some patients will significantly improve their comprehension of spoken words for both trained and untrained items following Listen-In treatment.

(3) At the individual level, patients who significantly improve their comprehension of spoken words for both trained and untrained items, will show a significant improvement in their phonological discrimination ability following treatment.

(4) Patients will not show changes in non-verbal auditory processing skills.

(5) Baseline demographic and behavioural factors will explain a proportion of the variance in treatment outcomes.

Word level treatment studies suggest item specific effects are the predominant response type amongst patients with aphasia, therefore the first hypothesis was that improvements would occur for treated items compared with matched, untreated items (Hypothesis 1). Two previous studies which investigate phonological-semantic treatment solely at the whole word level have reported generalisation to untreated items; however, these effects may have been due to generalised improvements unrelated to speech perception, and there were only four patients (Knollman-Porter et al., 2018; Raymer et al., 2006). Therefore a tentative hypothesis, given this paucity of previous research, is that speech comprehension training will improve speech perception leading to improvements across treated and untreated items (Hypothesis 2). The next hypothesis (Hypothesis 3) was that if patients do improve on untreated items, then this may be due to

improvements in phoneme discrimination, given that this aspect of processing is being targeted in Listen-In. Finally, it was hypothesised that patients would show variability in response to treatment, in line with a wealth of previous research, and that baseline demographic and behavioural variables may be able to account for some of the variance in treatment outcomes.

5.2.1 Experimental Procedures

5.2.1.1 Participants

Data from thirty-five patients were included in the present analysis. Demographic and baseline data can be seen in Methods.

5.2.1.2 Design

A randomised, repeated measures cross-over design was used, with five, evenly-spaced testing time points (T1-T5) at 12-week intervals. The design of the study is described in detail in Methods.

5.2.1.3 Data preprocessing

Listen-In treatment data

Dose (rounded to hours) was the total time spent on the therapy component of Listen-In over the therapy block. Challenge exposure was the number of individual challenges which the patient completed over the 12-week therapy block. Dose and challenge data were recorded automatically by the app and entered into Excel manually for each patient.

Baseline and repeated measures outcome measures

Behavioural scores were extracted for all outcome measures across the five time points. Data were entered into Excel, and statistical analyses were carried out using SPSS24.

5.2.1.4 Statistical analyses

Aim 1: To investigate whether changes in performance were specific to Listen-In treatment, repeated measures ANOVAs compared change over Listen-In treatment with a 12-week control period (standard care), in line with the cross-over study design.

Aim 2: To ascertain individual treatment effects, a McNemar test was carried out independently on each patient. Two separate analyses were conducted for each patient, one for treated items, and one for untreated items. This repeated-measures test investigated whether patients made significantly more incorrect-to-correct responses from pre to post treatment, compared with correct-to-incorrect responses.

Aim 3: Simple correlations were run to investigate whether treatment dose (hours) or baseline severity were associated with response to treatment. To investigate whether a combination of baseline variables could explain some of the variance in treatment outcomes, automatic linear modelling (ALM) in SPSS24 was then conducted.

5.2.2 Results

Definitions of specific terms used in this section:

Trained Challenge Item: 110 items trained during Listen-in treatment and tested in the ACT in the same linguistic construction, tested at each time point (T1-T5)

e.g. ACT: "The computer"

Listen-In: "The computer"

Trained Lexical Item: 110 items trained during Listen-in treatment, where the target word is a trained item, but the linguistic construction in Listen-In is different from the one tested in the ACT

ACT: "The computer"

Listen-In: "Computer", "He likes the computer", "The computer is new"

Core therapy challenge: all other core therapy items trained during Listen-In treatment, but not tested in the ACT

5.2.2.1 Listen-In usage data

5.2.2.1.1 Treatment dose

Figures 5-9 and 5-10 display usage data for the Listen-In treatment app, for each patient. Over the treatment block, patients spent an average of 85 hours (SD=31) on the therapy component, and 24 hours (SD=27) on the gaming component. Patients completed a large number of individual therapy challenges (a single word-to-picture matching challenge) over the treatment period, completing an average of 3098 (SD=999) Trained Challenge Items, 1633 (SD=1071) Trained Lexical Items, and a further 22217 (SD=11068) Core Therapy Items. As expected, hours spent on therapy (dose) correlated highly with the number of challenges patients completed: Core Therapy Items (r=.87, p<.001); Trained Challenge Items (r=.91, p<.001); and total Trained Lexical Items (r=.80, p<.001). Due to the correlation between hours spent on therapy and number of challenge completed, only dose was used as the independent variable to represent Listen-In usage in all further analyses.

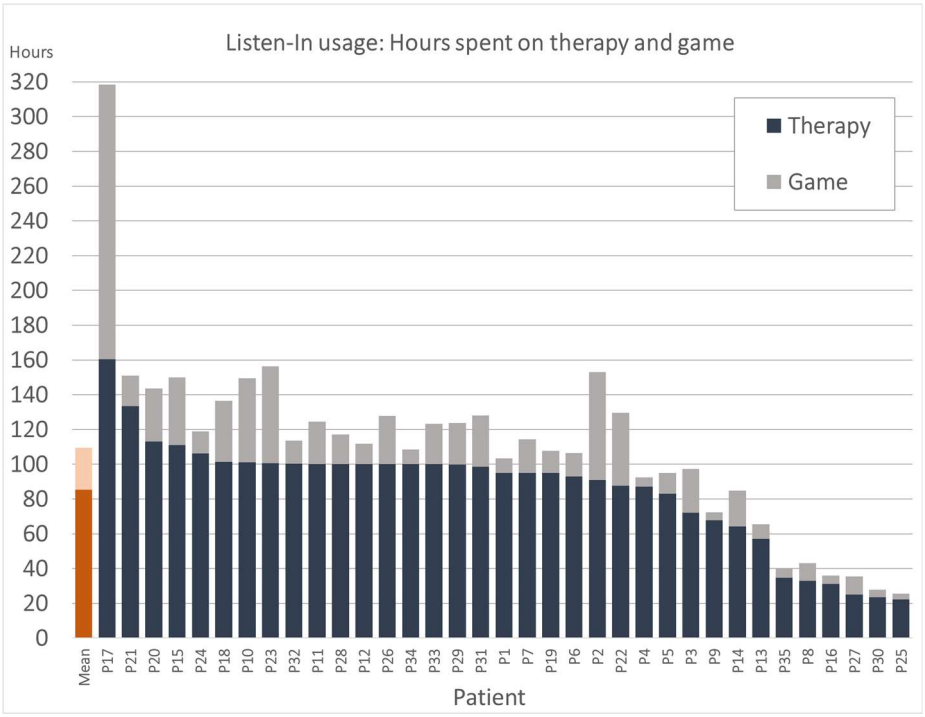


Figure 5-9 Time spent on Listen-In across patients (N=35) over the 12-week treatment block. Blue bars show time spent on therapy challenges, grey bars show time spent on game. Orange bar displays the mean. Patients sorted in order of therapy hours completed.

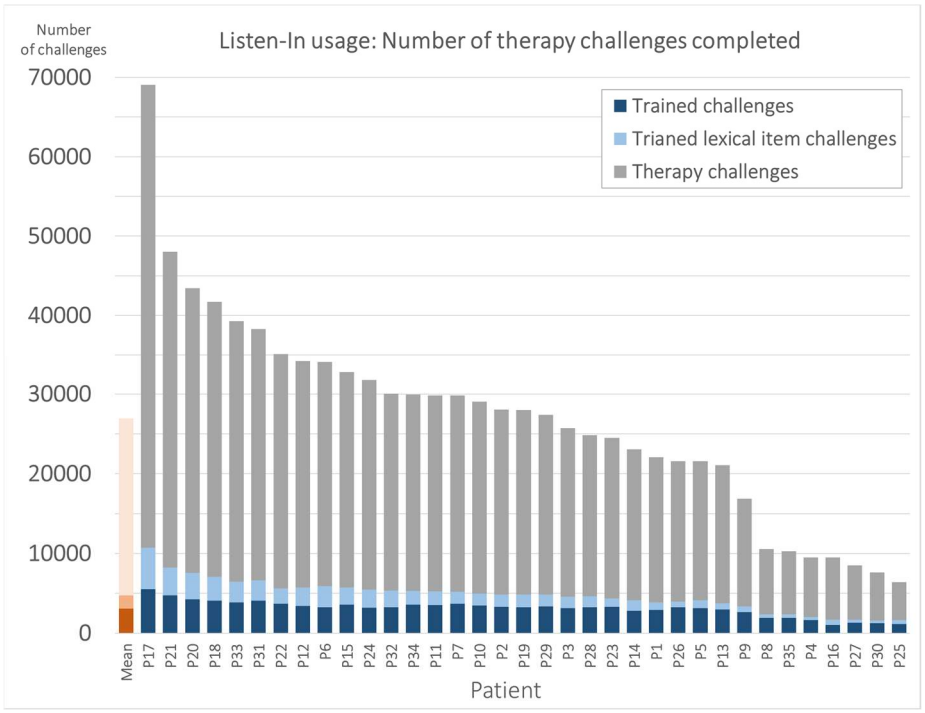


Figure 5-10 Number of completed challenges on Listen-In across patients (N=35) over the 12-week treatment block. Dark blue bars show exposure to Trained Challenge Items; light blue bars show exposure to trained lexical; and grey bars show exposure to Core

Therapy Items. Orange bar represents the mean. Patients sorted in order of total challenges completed.

5.2.2.1.2 Exposure to treated items

Mean exposure to individual Trained Challenge Items was 28 (SD=9), and mean exposure to trained lexical challenges was 15 (SD=10). Patients varied in pattern of exposure over time. Figure 5-11 displays exposure counts for treated items across two patients, to illustrate the variability in exposure to treatment stimuli over the course of therapy. Data show that P18 completed 2780 blocks over treatment, and P3 completed 1713 blocks over treatment. (Note these graphs to not include data from the remaining Core Therapy Items which were not tested items, so only blocks which contained a trained item are shown). These graphs show that:

- P18 and P3 had consistent exposure of Trained Challenge Items over treatment, in line with the algorithm which inserted treated item blocks at regular intervals.
- However, for some Trained Lexical Items, there was less consistency in how frequently the items were presented across treatment blocks, for the duration of therapy.
    - Data for P18 shows that approximately 80 Trained Lexical Items were exposed fairly evenly across treatment blocks. However, for approximately 30 Trained Lexical Items, exposure only occurred towards the beginning of treatment. This resulted in the patient being trained mainly on the Trained Challenge Item for 30 words these (the exact word/phrase/sentence tested in the ACT) for the duration of their therapy.
    - Data for P3 shows a broadly similar pattern, but with a greater proportion of Trained Lexical Items (approximately 50) showing inconsistent exposure. In other words, these items were mainly given at the beginning of therapy. This resulted in the patient being trained mainly on the Trained Challenge Item for these 50 words (the exact word/phrase/sentence tested in the ACT) for the duration of therapy.

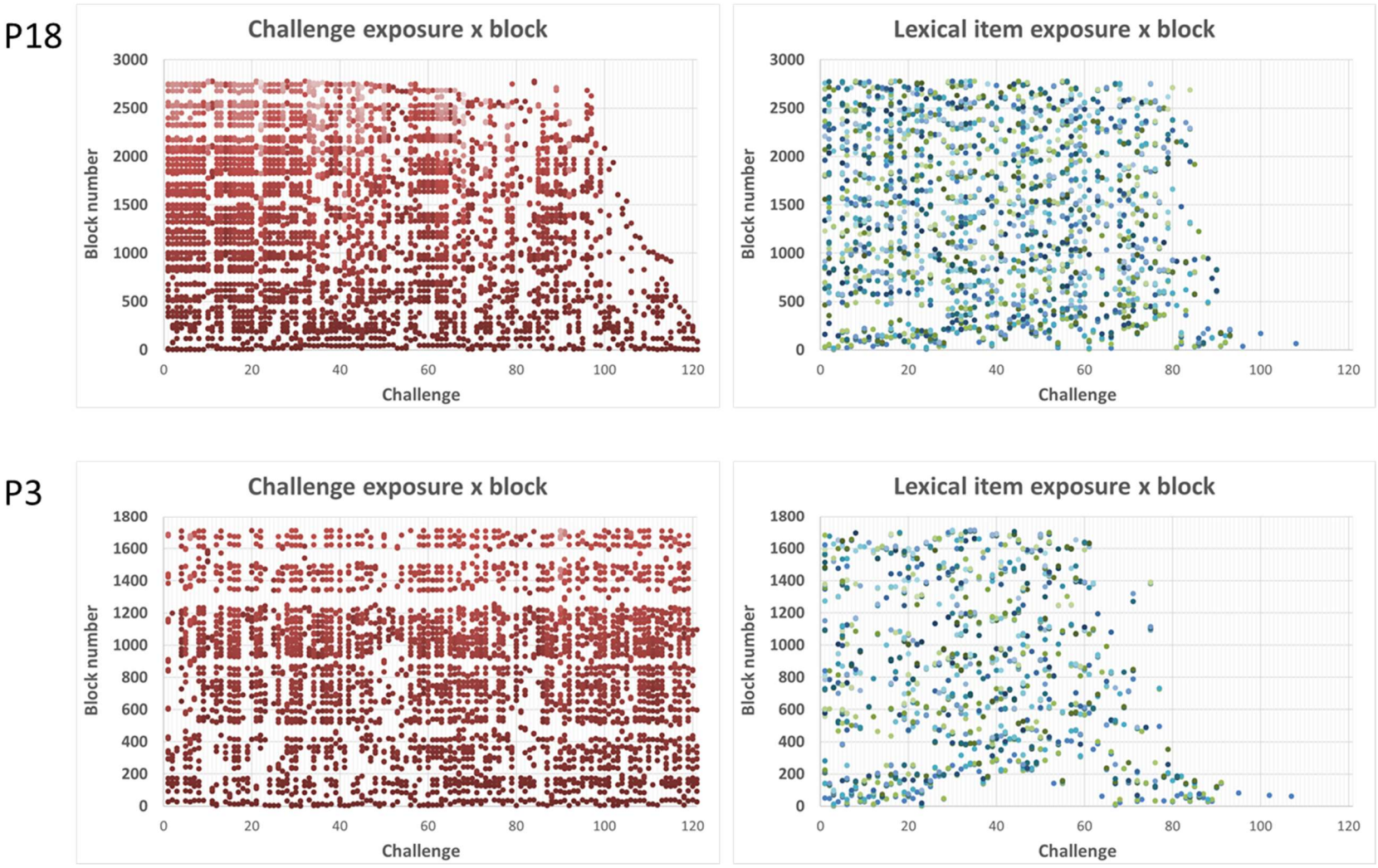


Figure 5-11 Graphs showing exposure to Trained Challenge Items (red) and Trained Lexical Items (blue) over the Listen-In treatment block, for two patients. One dot represents one single item. Items are sorted in order of range (maximum to minimum) of block exposure separately for each graph.

5.2.2.2 Group results

5.2.2.2.1 Cross-over group comparison for key baseline measures

There were no significant differences in Listen-In usage for Group 1 versus Group 2 on time spent on therapy (t(33)=1.69, p=.10) or the game (t(33)=.61, p=.55). A between groups comparison also indicated there were no significant differences between Group 1 and Group 2 on age (t(33)=-.40, p=.70) and time since stroke (t(33)=-.33, p=.75). In line with the minimisation randomisation method used, there were no significant differences in performance for Group 1 versus Group 2 on Spoken Words (t(33)=-1.25, p=.22) or Spoken Sentences (t(33)=-.71, p=.48). At baseline (T1), Group 1 performed significantly worse than Group 2 on the ACT (t(33)=-2.52, p=.02).

5.2.2.2.2 Baseline performance: Auditory Comprehension Test

Patients mean score at baseline was 53% (SD=12.97), many standard deviations below a group of 22 healthy age matched controls (M=95.25%, SD=2.27) (t(37)=18.81, p<.001) (equal variances not assumed). There was a small but significant improvement of 2.33% from T1 to T2 across all patients (t(34)=-2.38, *p*=.02). One participant demonstrated unreliable baseline performance (18% improvement from T1 to T2, an extreme outlier); on removal, a small but non-significant improvement of 1.87% over baseline time points remained across the group (t(33)=-2.09, *p*=.05), demonstrating overall stable baseline performance with this outlying patient removed.

5.2.2.2.3 Baseline performance: Cognitive and linguistic measures

Performance on further outcome measures can be seen in Table 5-2, alongside cut off criteria for each assessment where available. (Distributions of patients performance across measures are provided in Appendix 3).

Linguistic

All patients scored below the aphasia cut-off for speech comprehension and production on the Comprehensive Aphasia Test, indicating no patients had deficits confined to the input or output modality. On the BPVS, nineteen scored over 120 indicating these patients were able to comprehend a relatively large number of single words within this test. In contrast, the remaining sixteen patients showed variable performance indicating significant impairment in comprehending single words (without semantic or phonological distractors).

Auditory processing

In identification of environment sounds (ENVASA), patients showed a mean accuracy of 55% (SD=26), indicating significant errors in identification of environmental sounds. Participants showed a significant increase in performance from T1 to T2 (T2=64%, SD=25) suggesting some of this variability may have been due to task demands. The majority of patients required several practice trials before the task was sufficiently understood. Patients showed a mean of 54% (SD=17) on the PDT, with a normal distribution, suggesting this task captured a range of impairments in phoneme discrimination. Thirteen patients scores were not significantly different from change suggesting at least 13 patients had moderate-severe phoneme discrimination impairments, or difficulty with other task demands.

Cognitive and semantic

All patients scored below age-matched healthy controls on the CATTELL, indicating nearly all patients had non-verbal cognitive deficits (M=12, SD=4). On the picture matching task (SAT), around a third of patients made considerable errors, indicating many patients had semantic level impairments.

Baseline Aphasia profiles

All patients had aphasic impairments across speech comprehension, naming, and repetition, as measured by composite scores on the Comprehensive Aphasia Test (for figures showing patients profiles, see Appendix 4). Table 5-2 shows patients broad aphasia types based on scores on the CAT. Wernicke's aphasia (WA) is characterised by severely impaired single-word comprehension and repetition, and speech which is fluent but disordered. Eight patients in the present study presented with fluent speech output, with impairments in both repetition

and naming (Fluent). The remaining 26 patients presented with non-fluent speech, or no speech output, and impairments across comprehension and spoken output, corresponding to global aphasia, although many showed variable performance across the different domains. Twelve patients showed relatively equal impairments across all subtests (Global). Nine patients showed slightly poorer comprehension relative to speech output suggesting a particular impairment in speech comprehension (Global – comprehension -). Four patients showed better comprehension and poorer speech output, one showed marked impairment in repetition, and one showed better comprehension and repetition in light of poor naming (Global – comprehension +).

(P35 had markedly less impaired repetition almost within the normal range, with poor comprehension. Of note was that this patient had an additional right hemisphere lesion, performed the poorest out of the group on SART-errors, demonstrated inconsistent performance on the ACT across timepoints, and presented with other behavioural signs which suggest difficulty regulating responses (e.g. pressing button at the right time, repetitive verbal output) suggesting further cognitive impairments).

Table 5-2. Baseline performance across key language and cognitive measures

ID	Group	Speech Comprehension		Speech Production		ACT (%)	BPVS (%)	PDT (%)	ENVASA (%)	CATTELL (%)	SAT (%)
		Words (/30)	Sentences (/32)	Naming (/48)	Repetition (/74)						
P1	1	**11**	**10**	**9**	**11**	**41**	18	**40**	19	**46**	50
P2	1	**19**	**11**	**42**	**27**	**43**	84	**35**	10	50	80
P3	1	**23**	**10**	**5**	**14**	**55**	46	50	85	**23**	90
P4	1	**23**	**15**	**7**	**0**	**65**	82	50	59	62	97
P5	1	**20**	**9**	**0**	**16**	**32**	51	57	41	**23**	40
P6	1	**15**	**14**	**27**	**45**	**42**	73	69	29	**35**	57
P7	1	**10**	**8**	**0**	**0**	**42**	44	29	63	**38**	90
P8	1	**17**	**7**	**15**	**33**	**55**	45	83	95	**46**	67
P9	1	**8**	**11**	**2**	**15**	**45**	86	**36**	58	**42**	97
P10	1	**19**	**10**	**11**	**40**	**54**	81	67	0	**42**	97
P11	1	**12**	**11**	**32**	**36**	**51**	57	**43**	21	**15**	43
P12	1	**22**	**8**	**6**	**20**	**50**	53	52	44	**42**	87
P13	1	**9**	**12**	**39**	**28**	**47**	51	45	75	62	90
P14	1	**19**	**10**	**38**	**37**	**40**	54	52	58	**27**	87
P15	1	**15**	**2**	**16**	**22**	**57**	84	**43**	64	65	83
P16	1	**22**	**8**	**2**	**0**	**57**	64	**40**	74	50	83
P17	1	**14**	**4**	**2**	**8**	**31**	37	62	61	**31**	23
P18	1	**20**	**5**	**3**	**2**	**56**	78	**43**	84	58	93

*Table 5-2 Baseline performance across key language and cognitive measures. *P25 and P33 withdrew at T5 due to illness. P35 completed partial assessments at T5 as they requested a shorter testing session. SD=standard deviation. ACT=Auditory Comprehension Test; BPVS=British Picture Vocabulary Scale; PDT-Phoneme Discrimination Test; ENVASA= Environmental Sounds Test; CATTELL=Culture Fair Intelligence Test; SAT=Semantic Association Test.*

ID	Group	Speech Comprehension		Speech Production		ACT	BPVS	PDT	ENVASA	CATTELL	SAT
		Words (/30)	Sentences (/32)	Naming (/48)	Repetition (/74)	(%)	(%)	(%)	(%)	(%)	(%)
P19	2	**22**	**12**	**38**	**0**	**62**	73	65	63	**27**	77
P20	2	**23**	**3**	**36**	**30**	**63**	79	67	69	58	90
P21	2	**25**	**4**	**2**	**1**	**57**	76	67	53	**19**	87
P22	2	**0**	**3**	**0**	**0**	**28**	5	**40**	23	**27**	47
P23	2	**16**	**6**	**0**	**0**	**57**	79	62	50	**42**	90
P24	2	**21**	**15**	**55**	**15**	**62**	86	48	45	**31**	100
P25*	2	**12**	**11**	**52**	**38**	**38**	26	**33**	69	**38**	90
P26	2	**23**	**9**	**0**	**3**	**64**	78	74	83	65	97
P27	2	**9**	**4**	**0**	**2**	**39**	42	**43**	74	**50**	90
P28	2	**22**	**18**	**0**	**0**	**72**	88	**40**	71	62	90
P29	2	**30**	**12**	**51**	**52**	**79**	93	81	75	58	87
P30	2	**24**	**15**	**1**	**4**	**46**	38	74	93	**42**	77
P31	2	**25**	**13**	**0**	**1**	**69**	88	90	80	58	90
P32	2	**24**	**18**	**40**	**28**	**49**	42	**40**	9	**35**	77
P33*	2	**17**	**9**	**12**	**36**	**56**	72	48	6	58	93
P34	2	**14**	**8**	**0**	**0**	**67**	85	**40**	50	54	97
P35*	2	**19**	**13**	**40**	**62**	**83**	88	86	83	65	90
Mean:		**18**	**10**	**17**	**18**	**53**	**64**	**54**	**55**	**44**	**81**
Group 1:		17	9	14	20	48	60	50	52	42	75
Group 2:		19	10	19	16	58	67	59	59	46	86

Bold indicates scores as follows:

- Spoken Words=25; Spoken Sentences=27; Naming=69; Repetition=67 (below aphasia cut off scores on the Comprehensive Aphasia Test).
- CATTELL: <50% = 1 SD below the mean based on a healthy control group (N=27; mean age= 56 years (SD=12);mean accuracy=66% (SD=14)).
- ACT: <93% = 1 SD below the mean (<91% 2 SD; <88% 3 SD)
- PDT: 24-43% indicates performance not significantly different from chance (p>.05)

5.2.2.2.4 Treatment effects: Auditory Comprehension Test

Performance on the ACT across time point is displayed in Figure 5-12, by group. In line with the crossover design of the study, a two-way ANOVA compared change in performance for treated and untreated items over treatment and standard care blocks. There were two within subject factors: 1) block (treatment block vs. standard care block); and item type (treated vs. untreated items); and one between subjects factor: group (Group 1 vs Group 2). A significant interaction between treatment block and item type was found ($F(1, 33)=39.16$, $p<.001$).

Treated items

Patients improved their comprehension of treated items over the treatment block by an average of 10.55% (SD=9.43), and this effect size was large, as indicated by a one sample t-test ($t(34)=6.61$, $p<.001$, Cohen's $d=1.1$). Patients showed no significant change in performance for treated items over the standard care block (M=-.56, SD=8.73, $t(34)=-.38$, $p=.71$, Cohen's $d=-0.06$). A paired samples t-test compared change over the treatment block with change over standard care, in line with the cross-over study design. This showed that patients made significantly greater improvements over the Listen-In treatment block versus standard care block, with a mean difference of 11.10% (SD=16) ($t(34)=4.09$, $p<.001$).

Untreated items

Patients showed a small decline in performance for untreated items over the treatment block of -.93% (SD=9.64), but improved their comprehension of untreated items over the standard care block by 4.96% (SD=7.75). The Kolmogorov-Smirnov statistic (p=.005) indicated a non-normal distribution, due to one extreme outlying patient, whose performance for untreated items got worse over the treatment block by -39%. A non-parametric Wilcoxen signed-ranks test was therefore used to compare change in untreated item performance over the treatment and standard care blocks. This revealed that patients improved their comprehension of untreated items significantly more over

the standard-care block, compared to the Listen-In treatment block (Z=450, *p*=.03). For Group 1, this was driven by an improvement of 5.60% in the 12-week period following treatment (T4-T3) (see Table 5-3). For Group 2, this was driven by an improvement of 4.28% in the 12-week period prior to treatment (T3-T2). This indicates that these medium-sized improvements were not specific to Listen-In treatment.

Performance for treated and untreated items in maintenance blocks

Change in performance for treated items over the maintenance blocks can be seen in Table 5-3 for each group. Maintenance effects are reported separately for each group due to difference in length of maintenance periods.

Treated items

For Group 1, there was a non-significant decline of -3.59% in speech comprehension in the 12-week period immediately following treatment (T3-T4) (t(17)=-2.01, p=.06). In the subsequent maintenance block there was no significant change in treated item performance (T4-T5) (t(17)=.39, p=.70). For Group 2, there was no significant change in performance in the 12-week period following treatment (T4-T5) (t(14)=-.14, p=.89). In summary, there were no significant declines in treated item performance at 12 and 24 weeks post treatment, indicating maintenance of treatment effects.

Untreated items

Group 1 showed a significant improvement for untreated items in the 12-week period immediately following treatment (T4-T3) (t(17)=3.01, p=.008). In the subsequent maintenance block there was no significant change in untreated item performance (T5-T4) (t(17)=1.05, p=.31). Group 2 showed a small improvement in performance for untreated items in the 12-week period following treatment, but this was not significant (t(14)=1.32, p=.21).

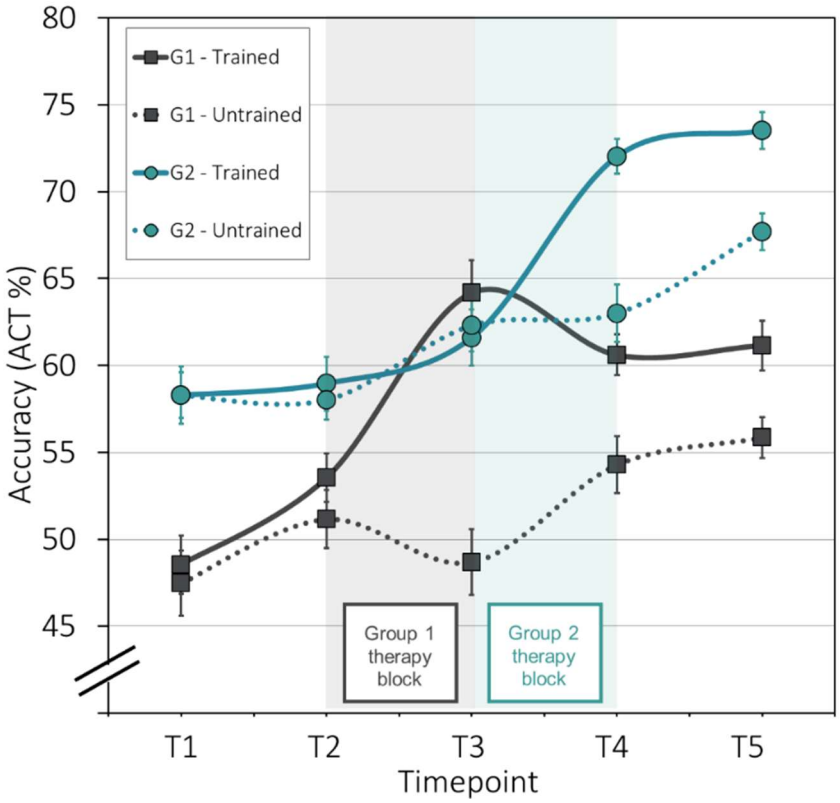


Figure 5-12 Performance on the Auditory Comprehension Test, for treated and untreated items, by group. G1=Group 1; G2=Group 2. Treated = items treated during Listen-In treatment block; Untreated = items not treated during Listen-In treatment block. Error bars are within subject standard error of the mean (for Group 2, two patients were excluded from error calculations due to no T5 data).

	Treated items				Untreated items			
	T1 – T2	T2 – T3	T3 – T4	T4 – T5	T1 – T2	T2 – T3	T3 – T4	T4 – T5
Group 1								
% change	5.00*	10.65**	-3.59	0.56	3.68*	-2.47	5.60*	1.57
95% CI	2.59	4.52	3.57	2.85	2.88	4.73	4.07	3.06
Cohens *d*	0.8	1.0	-0.5	0.1	0.5	-0.2	0.7	0.3
Group 2								
% change	0.67	2.65	10.43**	-0.12	-0.27	4.28*	0.70	2.12
95% CI	3.26	4.63	3.51	1.63	3.54	40.77	3.42	3.09
Cohens *d*	0.1	0.3	1.2	-0.04	-0.04	0.6	0.1	0.3

*Table 5-3 Unstandardised and standardised effect sizes for change in performance on the Auditory Comprehension Test, by group, across trial periods. Shaded areas represent Listen-In treatment block. CI=confidence interval based on within subject standard error of the mean. Group 2: T4-T5 data from N=15, as two participants have no data for T5. Independent samples t-tests on change scores: *p<.05, **p<.001.*

5.2.2.2.5 Treatment effects: Spoken Words and Sentences (Comprehensive Aphasia Test)

Change scores across time points are displayed in Table 5-4 by group. Two repeated measures ANOVAs compared change in performance over treatment and standard care blocks for CAT-words and CAT-sentences. For both analyses, there was one within subject factor: change over block (treatment block vs. standard care block); and one between subjects factor: group (Group 1 vs Group 2). For CAT-words, there was no significant main effect of block ($F(1, 33)=1.87$, $p=.18$), and no significant between subjects effect of group ($F(1, 33)=.00$, $p=.99$). For CAT-sentences, there was also no significant main effect of block ($F(1, 33)=.52$, $p=.48$), and no significant between subjects effect of group ($F(1, 33)=.99$, $p=.33$). Paired samples t-tests showed no significant difference in performance change over treatment compared to standard care blocks for both CAT-words ($t(34)=1.40$, $p=.17$) and CAT-sentences ($t(34)=-.74$, $p=.46$).

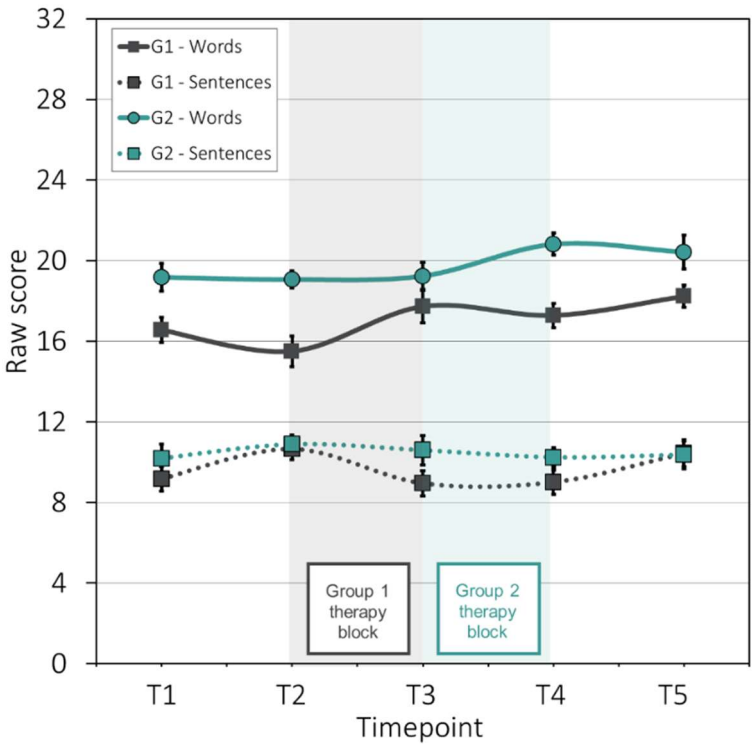


Figure 5-13 Performance on Spoken Word and Spoken Sentence subtests from the Comprehensive Aphasia Test, by group. G1=Group 1; G2=Group 2. Error bars are within subject standard error of the mean (for Group 2, two patients were excluded from error calculations due to no T5 data).

	CAT Spoken Words				CAT Spoken Sentences			
	T1 – T2	T2 – T3	T3 – T4	T4 – T5	T1 – T2	T2 – T3	T3 – T4	T4 – T5
Group 1								
Mean change (raw score)	-1.06	2.22	-0.44	0.94	1.50	-1.72	0.06	1.44
95% CI	2.15	2.67	2.51	1.82	1.66	1.92	1.95	1.90
Cohens *d*	-0.2	0.4	-0.1	0.21	0.4	-0.4	0.2	0.32
Group 2								
Mean change (raw score)	-0.12	0.18	1.59	-0.07	0.71	-0.29	-0.35	0.86
95% CI	1.78	1.71	2.21	2.19	1.92	2.17	2.23	1.72
Cohens *d*	-0.03	0.1	0.4	-0.01	0.2	-0.1	-0.1	0.25

Table 5-4 Mean change in performance (raw scores) and standardised effect sizes for performance on Spoken Word (max 30) and Spoken Sentence (max 32) subtests on the Comprehensive Aphasia Test, by group, across trial periods. Shaded areas represent Listen-In treatment blocks. CI=confidence interval, based on within subject standard error of the mean. Cohen's d effect sizes are one sample t-tests. Group 2: confidence intervals based on N=14 due to no subtest data for three participants.

5.2.2.2.6 Treatment effects: Cognitive and linguistic measures

Four repeated measures ANOVAs compared change in performance over treatment and standard care blocks for further secondary outcome measures. For all analyses, there was one within subject factor: 1) change over block (treatment block vs. standard care block); and one between subjects factor: group (G1, G2). There were no was significant main effects of block for all measures. These included: PDT ($F(1, 33)=1.23$, $p=.28$); BPVS ($F(1, 33)=2.70$, $p=.11$); ENVASA ($F(1, 33)=.001$, $p=.97$); and SART-errors ($F(1, 33)=1.01$, $p=.32$). There were also no significant interactions between block and group, or main effect of group, for all outcome measures.

5.2.2.2.7 Treatment effects: Self-reported outcomes

Self-report ratings were ordinal data based on a Likert scale response, therefore three Wilcoxen signed ranks tests were run to compare changes in self-ratings over treatment compared to standard care, on three subtests. These revealed no significant changes in self-ratings for the expressive subtest ($Z=-.15$, $p=.88$), receptive subtest, ($Z=-.22$, $p=.83$), and daily tasks subtest ($Z=-.08$, $p=.94$).

5.2.2.3 Individual treatment effects

For the following analyses, McNemar's test was calculated individually for each patient, with the alpha set to p<.05.

5.2.2.3.1 Treated items

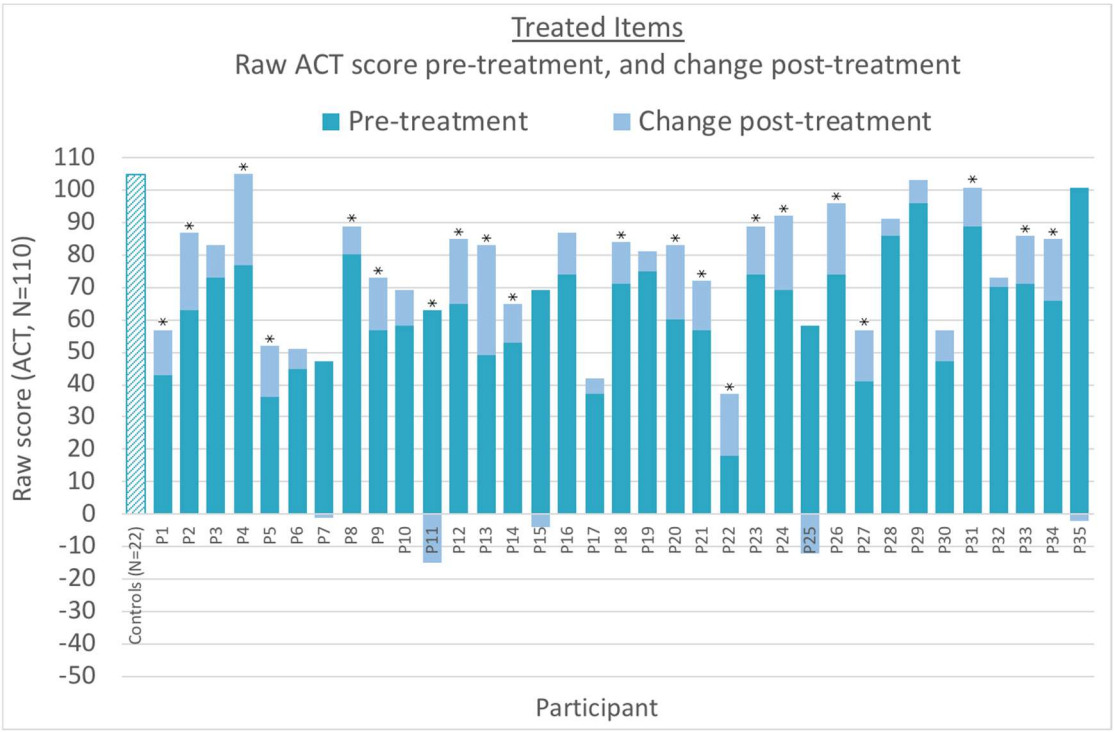
Performance across test items

McNemar's test indicated that twenty patients made proportionally more correct than incorrect responses from pre to post treatment (p<.05), with overall improvements of between 8% and 31% (see Figure 5-14). Ten patients demonstrated small improvements (3 to 12%), but the differences between correct-to-incorrect and incorrect-to-correct responses from pre to post treatment were not significant. Five patients showed declines post treatment (-1 to -15%), and this difference was significant for one participant. In summary, twenty patients comprehended significantly more treated items post treatment, whilst ten patients showed small but non-significant gains.

5.2.2.3.2 Untreated items

For untreated items, McNemar's test indicated that two patients made significantly more incorrect-to-correct, than correct-to-incorrect, responses from pre to post treatment (p<.05) (10% and 13%) (see Figure 5-14). Two patients made significantly more correct-to-incorrect than incorrect-to-correct responses (-20% and -39%). The remaining thirty-two patients showed small and non-significant improvements and declines (12% to -12%) in performance from pre to post treatment. In summary, two patients comprehended significantly more untreated items post treatment.

(A)



(B)

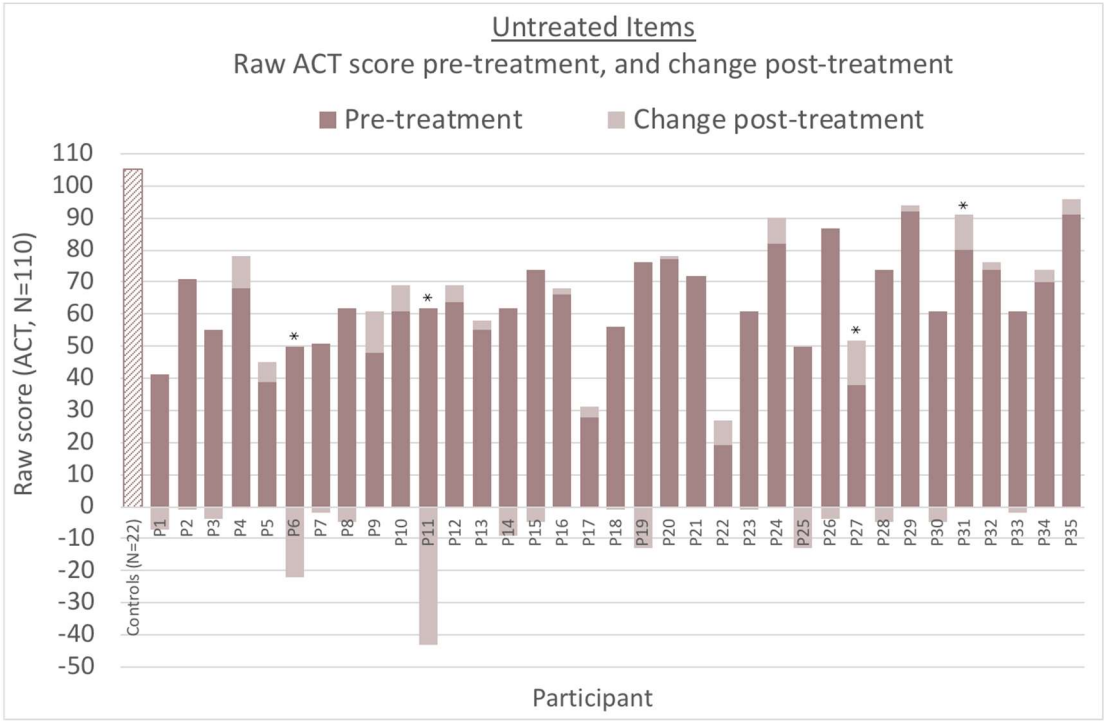


*Figure 5-14 Number of items correct on the Auditory Comprehension Test (ACT) at pre- and post- treatment timepoints for treated (A) and untreated (B) items. Controls=mean accuracy for 22 age matched controls across all items, halved to provide a comparison (N=220/2) *McNemar's test indicates a significant difference between incorrect-to-correct and correct-to-incorrect responses from pre to post treatment (p<.05).*

5.2.2.4 Explaining response to treatment

5.2.2.4.1 Relation between dose and initial severity and treatment outcomes

Simple correlations were run to investigate any positive correlation between the key dependent variable (percentage improvement on treated items) and two possible explanatory variables: time spent on Listen-In therapy challenges (dose), and baseline ACT performance (initial severity). Results revealed that dose ($r(35)=.08$, p=.64) and initial severity ($r(35)=-.006$, p=.98) bore no relation with treatment outcomes, as displayed in Figure 5-15.

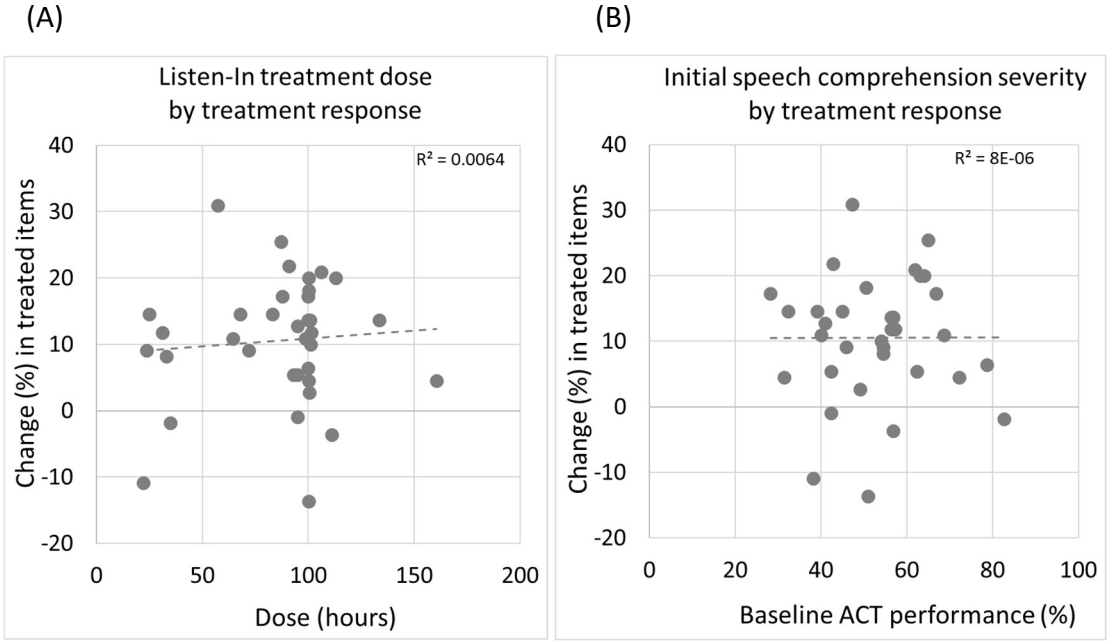


Figure 5-15 Scatterplots showing simple correlations between the dependent variable (% change on treated items) and two possible explanatory factors: (A) hours spent on Listen-In therapy; (C) baseline ACT performance.

5.2.2.4.2 Explanatory modelling - Automatic linear modelling (ALM)

To investigate whether a combination of baseline measures could explain some of the variance in treatment outcomes, regression analysis was conducted using automatic linear modelling (ALM) in SPSS24. The dependent variable was percentage improvement on treated items from pre to post treatment. There were fifteen predictor variables: age,

time since stroke (TSS), and baseline performance on the ACT, SAT, CATTELL, ENVASA, BPVS, PDT, SART-errors, CAT-words, CAT-sentences, CAT-repetition, CAT-naming-nouns, CAT-naming-verbs, and CAT-verbal-fluency. To allow all combinations of predictor variables to be modelled, the best subtests model selection method was selected. Automatic preparation of data was turned on. All other settings were kept as default.

The results identified a best model which included nine baseline variables (Figure 5-16). In order of predictor importance these were: CAT-repetition ($\beta = -0.43$, $p < .001$), ENVASA ($\beta = -0.22$, $p = 0.01$), CAT-verbal-fluency ($\beta = 1.46$, $p = 0.01$), PDT ($\beta = 0.27$, $p = 0.03$), ACT ($\beta = -0.37$, $p = 0.06$), SAT ($\beta = 0.52$, $p = .16$), BPVS ($\beta = .09$, $p = 0.16$), age ($\beta = -0.17$, $p = 0.26$), and CATTELL ($\beta = .50$, $p = 0.29$). Four variables showed a negative correlation with the dependent variable (Cat-repetition, ENVASA, ACT, age), indicating younger patients, and those with poorer scores on these measures, tended to show greater response to treatment. The remaining 5 variables showed a positive correlation (CAT-verbal-fluency, PDT, SAT, BPVS, CATTELL), with greater performance on these measures associated with greater treatment outcomes. The adjusted R^2 for the explanatory model was 0.36, indicating this combination of variables accounted for 36% of the variance in treatment outcomes. The model identified four participants who had outlying dependent variable scores (change in treated item performance). None of these values were close to 1 (Cook's Distance=.36, .23, .18, .17), indicating minimal influence on the best model, therefore these data points were not removed from analyses.

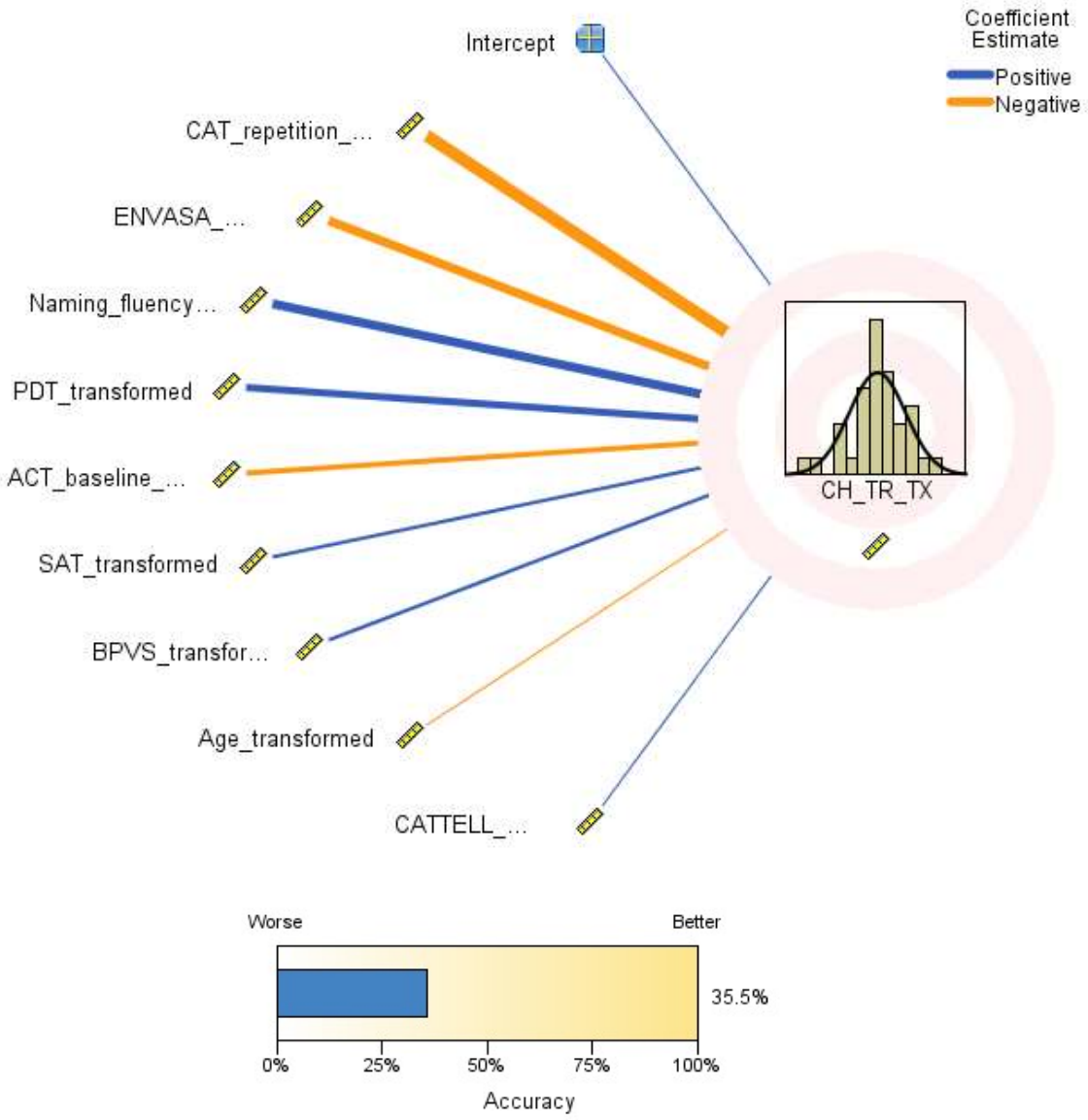


Figure 5-16 Demographic and baseline behavioural measures model. The diagram represents nine variables which, when combined, explain 36% of the variance in treatment outcomes. Orange lines=negative associations; blue lines=positive associations. The blue bar at the bottom represents the model accuracy (adjusted R^2).

5.2.3 Discussion

5.2.3.1.1 Summary of results

Self-administered tablet-based therapy significantly improved aphasic patients comprehension of spoken words. The largest change in speech comprehension was over the Listen-In treatment block, for both cross-over groups, indicating that improvements were specific to treatment. Patients improved, on average, by 11% for treated items, and treatment gains were maintained up to 24 weeks following cessation of therapy. These findings therefore support the first hypothesis, that gains would be demonstrated for items treated in therapy.

At an individual level, twenty patients showed significant improvements for treated items, whilst the majority of the remaining patients made small but non-significant improvements. These gains were specific to treated items; there were no significant improvements for untreated items at a group level, and at an individual level, only two patients showed significant gains in untreated items following treatment. These findings do not support the second hypothesis of untreated item effects for some patients. Instead, treatment effects were item specific across the majority of patients. For the two patients who made significant improvements, one participant (P27) improved their phoneme discrimination (on the PDT) over treatment but showed overall inconsistent performance over time, and the other participant (P31) demonstrated no improvements on this task, therefore Hypothesis 3 was not supported. At a group level, no significant changes over treatment were observed in environmental sound identification (ENVASA), supporting Hypothesis 4.

Simple correlations showed total dose and initial severity could not explain response to treatment. Further explanatory modelling of baseline variables indicated that a combination of nine variables best explained response to treatment, in support of Hypothesis 5. However, this only accounted for 36% of the variance in treatment outcomes, indicating the majority of variability in treatment outcomes could not be explained by behavioural or demographic patient characteristics. Repetition ability

showed the greatest importance within the model, more than any other variable, and indicated that individuals who entered treatment with poorer repletion skills tended to make greater improvements in comprehension of treated items.

5.2.3.1.2 Item specific treatment effects

Patients' comprehension of treated items significantly improved following Listen-In treatment, supporting Hypothesis 1. Hypothesis 2 stated that improvements may occur for treated and untreated items for some patients, but this was not generally supported, as only two patients showed significant improvements for untreated items. Patients also showed no significant changes in CAT-words performance over treatment compared to standard care, again supporting the specificity of treatment effects. Performance on CAT-sentences was stable across time points. This test assesses a variety of syntactic structures beyond single word comprehension, which were not treated in Listen-In, therefore this result unsurprising.

On average, patients improved by 11%, with improvements of up to 31%. The effect size was large (*d*=1.1). This is the first group study to report positive findings for this type of treatment, and therefore offers significant evidence for the efficacy of speech comprehension treatment in patients with chronic aphasia. The effect size compares favourably with those found for naming (Best et al., 2013), reading (Woodhead et al., 2018) and apraxia of speech (Varley et al., 2016) treatment studies in patients with chronic aphasia.

5.2.3.1.3 Comparison with previous treatment studies

Phonological based treatments

A number of case and group studies have reported improvements in patients' speech comprehension following phonological based treatments (Bastiaanse et al., 1993; Francis et al., 2001; Grayson et al., 1997; Morris et al., 1996; Tessier et al., 2007). Morris and colleagues (1996) describe a patient (JS) who improved on speech discrimination and

comprehension following treatment targeting only phonological discrimination. This treatment program included sublexical tasks, as well as spoken word-to-picture matching with phonological foils, and so also trained at the whole word level similar to Listen-In. Francis and colleagues (2001) describe an approach where the written modality was used in treatment to support access to comprehension, and report improvements in speech comprehension for treated items in their patient. The remaining patients differ in some respect from classic post-stroke chronic aphasic patients and so direct comparisons are difficult. Two patients were less than six months post traumatic brain injury and stroke, so spontaneous improvements may have affected findings (Bastiaanse et al., 1993; Grayson et al., 1997); and a further patient presented with a brain stem lesion and probable damage to the auditory pathway, rather than a more typical lesion to the fronto-temporo-parietal cortex (Tessier et al., 2007).

In contrast, other studies have reported poor or no treatment effects following phonological based treatments (Maneta et al., 2001; Prins et al., 1989; Woolf et al., 2014). Maneta and colleagues (2001) report one patient who showed no improvement following treatment with both sublexical and whole word tasks, which included written and spoken word to picture matching. In line with this single case, Woolf and colleagues (2004) found no treatment effects for 8 patients following 12 hours of therapy, based on a similar therapy regime.

One issue with treatment regimens described is that a range of different tasks have typically been implemented. For example, Maneta and colleagues (2001) used phoneme-to-grapheme matching, spoken-to-written word matching, and spoken word-to-picture matching, to target phoneme discrimination, and the total dose was 6 hours over 12 weeks. It may be that each task was not delivered in a sufficient dose and/or intensity to bring about change. In contrast, Listen-In used a single task, and provided a considerably larger dose (on average 85 hours). This is in keeping with several neuroscientific principles which have been proposed to support learning at the neural level: specificity, repetition, and intensity (Kleim & Jones, 2008). This dose is also in line with recent reviews which have found beneficial effects of high dose treatments of ~100 hours (Bhogal et al., 2003; Brady et al., 2016).

Patient performance on the PDT at baseline shows an average accuracy of 52%, with highly variable performance following a normal distribution. Aphasic performance has been shown to be variable on an adaptive version of this task (Robson, Keidel, et al., 2012), and so the present variability likely represents a range of discrimination abilities amongst patients. Of note is that 13 patients scored no better than chance. It is possible these patients had severe phoneme discrimination impairments; or, it may be that other aspects of the task were too difficult, as holding three sequences in short term memory, and then making a decision as to which is different, requires additional verbal auditory short term memory and executive skills. Given this range of phoneme discrimination ability, it is possible that some patients responded to the phonological aspects of task, accounting for treatment effects. However, any improvements in phonological processing do not appear to be related to generalised phoneme discrimination, as no changes were observed over treatment. It may be that phonological processing was improved, however, for the specific phonological-semantic networks of the items which were trained.

Semantic based treatments

A further possibility is that some patients responded to the semantic aspects of treatment. Previously, semantic based treatments targeting spoken word comprehension have reported positive findings (Francis et al., 2001; Morris & Franklin, 2012; Munro & Siyambalapitiya, 2017). Francis and colleagues' (2001) patient had word-meaning deafness, considered to be caused by impaired access between the word-form and meaning, rather than phonological impairment. Their patient improved for treated items, following a semantic treatment, with tasks such as reading definitions. Morris and Franklin (2012) carried out a similar word-picture verification task using semantic foils, with a total dose of ~6 and ~ 9 hours. Their two patients had good auditory/phonological processing, and impairments were located in lexical-semantics and conceptual-semantics. One patient improved their comprehension of treated and untreated items (~11%, in line with average Listen-In improvement), whilst the second patient made no improvements, despite similar baseline profiles. More recently, Munro and Siyambalapitiya (2017) describe a semantic feature analysis treatment with a globally aphasic patient, and reported improvements for trained words, and untrained words in the same semantic

category. Listen-In incorporated semantic foils, and many patients in Listen-In had impaired semantic processing, with around half of patients making considerable errors (>5) when matching semantically related pictures (SAT). It is therefore conceivable that patients with semantic level impairments may have benefitted from this aspect of the treatment program.

Phonological-semantic based treatments

To my knowledge, only one study has been reported which used a combined phonological-semantic therapy to target speech comprehension impairments, similar to the Listen-In task. In a feasibility study, Knollman-Porter and colleagues (2018) used a picture-word verification task, where the patient had to decide whether a spoken word matched an object. The spoken words were either semantic foils, or phonological foils, and so the target of treatment was the same as Listen-In. Two patients with severe auditory comprehension impairments each received 40 hours of treatment, and both patients demonstrated large and significant improvements across treated and untreated words. In a similar treatment task, but with the addition of written words, two patients with semantic impairments also showed 'some improvement' in comprehension of treated words following a computer based treatment (Raymer et al., 2006). Of note in these two studies is that both authors attribute improvements, in part, to generalised improvements in attention, awareness, or requests for repetition, due to the initial sharp increase in performance. However, both report treated item effects, showing at least some the improvements were specific to speech comprehension. For example, improvement between baseline and treatment phase accuracy was 52% for P1, and 71% for P2, much larger than the present study (Knollman-Porter et al., 2018). On the other hand, trained item performance was ~10-20% greater than untrained item performance for these two patients throughout the treatment phase, suggesting this reflects the true size of the treatment effect, in line with the effect size found for Listen-In patients (11%). The cross-over design of the present study demonstrates the effects found here are specific to treatment, and treated item effects also confirm improvements specific to speech comprehension.

In summary, previous treatment studies do not provide a clear picture, and methodological variations make assessing the efficacy of treatment a challenge, particularly in light of the overall small numbers of patients. These variations have included type of therapy tasks and target of treatment, dose, aphasia profile, study design, and outcome measures. The present findings build on the previous feasibility study (Knollman-Porter et al., 2018) by demonstrating that high-dose whole word treatment, combining semantic and phonological components, leads to significant improvements in speech comprehension for treated items.

There has been considerable debate in word retrieval treatments about whether phonological or semantic based treatments are most beneficial. Nickels (2002) suggests that combining both approaches allows processing along both dorsal (phonological) and ventral (semantic) routes. In the present study, it may be that both approaches contributed to treatment effects; however, it is not possible to assess their relative contribution based on the current data. Listen-In was designed to provide a broad treatment which targeted a range of levels of breakdown (phonological and semantic), to meet the needs of a typical caseload of patients. Following these positive findings, future studies may look to investigate the relative contribution of these components to treatment outcomes for different patients.

5.2.3.1.4 Explaining variability in response to treatment

Patients showed considerable variability in response to treatment. The explanatory model suggests that only a small amount of this variability can be explained by the baseline measures collected in this study. Within this model, a large number of variables were included. The variables which contributed to the model, in order of importance, were repetition ability (CAT-repetition), environmental sound discrimination (ENVASA), verbal fluency (part of CAT-naming), phoneme discrimination ability (PDT), spoken word comprehension (ACT), semantic association (SAT), vocabulary comprehension (BPVS), age, and a non-verbal measure of fluid intelligence (CATTELL). Given the large range of variables across a number of auditory, language and cognitive domains, it is difficult to identify a single underlying skill which explains how well patients responded to treatment.

The model identified repetition ability as the most important predictor, and this importance value was 1.8 times higher than that of the next most important predictor, environmental sound discrimination. This suggests that out of all the variables in the combined model, repetition ability contributed the most explanatory value to the model. However, it should be noted that many patients scored at floor level on repetition, and so there was little variability in performance, and this was not normally distributed across the group. Nevertheless, individuals who did show some repetition ability tended to be in the group of individuals who didn't significantly respond to treatment. In other words, patients who had poorer repetition skills tended to make treatment gains. One possibility is that individuals with poor repetition skills benefitted from the phonological component of treatment. Conversely, individuals with better repetition may not have benefitted as much from this type of treatment.

It is also interesting to note the top four behavioural predictors are all related to auditory and phonological processing, and are the only four predictors which independently show significant coefficients within the model. Taken together these suggest the following patient profile: poorer repetition, poorer environmental sound discrimination, better verbal fluency, and better phonological discrimination. Poor repetition and better verbal fluency are characteristic of Wernicke's type aphasia (Goodglass & Kaplan, 1972), and patients with speech comprehension impairments have been shown to have impaired non-verbal auditory processing and impaired phonological discrimination (Robson et al., 2013; Robson, Keidel, et al., 2012). The presence of these key measures as the top explanatory predictors in this model is therefore in line with factors known to be associated with speech comprehension impairments, lending validity to the hypothesis that these measures account for some of the variability in Listen-In outcomes. In other words, it may be that patterns of auditory and phonological processing are contributing a small amount of variability to how well patients respond to therapy, which ultimately suggests that patients with particular profiles may be more suited to this particular treatment. A further fine-grained analysis of auditory and phonological test performance may be warranted to investigate this possibility further. An alternative possibility, taken together with the collection of further variables, is that these simply indicate that more severely aphasic patients tended to make more gains. However, the presence of positive

and negative associations tends to support the hypothesis that it is particular patterns of performance which contribute to the model's explanatory power.

Overall, this model does not provide a clear picture of which patients are most likely to respond to treatment. Furthermore, even with this range of variables combined, the total explanatory power of this model is low. This is a clinically significant question, as patients devote considerable time and energy to high dose treatments, and so providing information on likelihood to respond is important for directing patients to therapies which are likely to improve their language outcomes. A future direction may be to include structural imaging data, to produce a combined model which may account for more variability. Aguilar and colleagues (2018) demonstrate this method in a group of patients with central alexia, who undertook a reading treatment. The authors found that a combined model accounted for 94% of the variability in treatment outcomes in their patients, which included: age, reading accuracy, reading comprehension, written semantic matching, and proportional damage to a number of brain regions in the LH. Crucially, the addition of lesion data significantly improved the explanatory power of the model, suggesting patterns of lesion damage contributed significantly to treatment outcomes. Aguilar and colleagues (2018) took their analysis a step further, by conducting out-of-sample analysis which attempted to predict performance for individual patients excluded from the model. This showed considerably less predictive power (23%). This discrepancy highlights the danger of over-fitting models to particular research samples; in the present sample, the poor explanatory model suggests a predictive model would not be suitable. However, the future inclusion of lesion data may improve the explanatory power of this model.

5.2.3.1.5 Generalisation of treatment effects

One hypothesis was that patients may show generalised improvements across all items (Hypothesis 2). The theoretical rationale was that training may improve phonological discrimination skills, which may then facilitate speech comprehension. This was not observed, as only two patients showed significant improvements for untreated items. Improvements were specific to word level comprehension, as two measures of auditory

processing showed no significant changes (non-word phoneme discrimination (PDT) and environmental sound discrimination (ENVASA)). These findings suggest that therapy effects were specific to the phonological and semantic networks which were engaged during spoken word comprehension. This may be through strengthened mappings between specific word forms and their semantic representations. In the cognitive neuropsychological model of speech comprehension, this relates to mappings between the phonological input lexicon and the semantic system (Whitworth et al., 2014). In addition to strengthening of mappings, it is feasible that training targeted the modules themselves. For example, someone with a greater phonological deficit may benefit via strengthening of the phonological representation itself (phonological input lexicon), while a patient with semantic level impairments may benefit by differentiating between semantic coordinates for a particular target word (e.g. dog/wolf). The present study did not set out to investigate between these possibilities, and it is likely that a combination of mechanisms contributed to treatment effects for different patients, given the heterogeneity of aphasic impairments in the present sample.

Generalisation to untrained words is a form of within-level generalisation, where improvements occur for untrained items which are at the same linguistic level as the focus of treatment (e.g. word level) (Webster, Whitworth, & Morris, 2015). Due to the paucity of treatment studies in speech comprehension, it was not clear whether this type of whole word treatment (Listen-In) would lead to generalised improvements. A few previous studies have reported generalisation to untreated items (Bastiaanse et al., 1993; Grayson et al., 1997; Knollman-Porter et al., 2018; Morris & Franklin, 2012; Tessier et al., 2007; Woodhead et al., 2017). However, they do not paint a clear picture: one patient was only two months post traumatic brain injury so generalisation may have related to general improvements (Bastiaanse et al., 1993); one improved following a phonological based treatment (Tessier et al., 2007); one improved with a semantic based treatment (Morris & Franklin, 2012); and in a group study, improvements occurred for spoken words which were not the target of treatment, but these were small and the study was conflated with the addition of a drug treatment (Woodhead et al., 2017). As previously described, Knollman-Porter and colleagues (2018) report improvements in untrained items in two patients with severe auditory comprehension impairments. However, it is questionable

whether improvements were related to speech comprehension or alternative processes, such as requests for repetition. In summary, there is too little evidence to assess generalisation in speech comprehension therapy. The present findings add to the limited evidence base by suggesting that traditional spoken word-to-picture matching facilitates item specific improvements, as least in the present sample. However, future studies are required to systematically investigate how manipulation of task parameters may affect generalisation.

In word retrieval treatments, which have a much larger evidence base than speech comprehension treatments, a review by Nickels (2002) found 13 out of 19 patients generalised to untreated items; however, treatments were strategy based and so the focus was different to Listen-In. Webster and colleagues (2015) reviewed studies which focused on remediating word retrieval impairments, the same approach as in Listen-In, and found few patients (21/69) showed within-level generalisation. In a review of 23 technology based anomia treatment studies, Lavoie, Macoir and Bier (2017) also report mixed findings. The present item specific treatment effects are therefore in line with findings from the naming literature.

Another finding Webster and colleagues (2015) note is that patients who do show within-level generalisation tend to have good semantics and poor phonological encoding. The two patients in Listen-In who did show possible generalisation to untrained items had differing aphasia profiles: both performed well on a non-verbal semantic association task, one patient had global aphasic deficits which included poor phoneme discrimination, and the other patient had milder aphasic deficits, mainly at the sentence level. No clear picture therefore emerges from these two patients.

A further question which arises is whether patients generalised from the specific item (I.E. target picture) to the conceptual object. For example, it is possible that treatment strengthened mappings between “dog” and the picture of a dog within the Listen-In app, but that this did not transfer to real objects beyond the task. Conversely, the most desirable outcome would be strengthening of this mapping with the underlying concept of “dog” which then transfers to other settings beyond the therapy task itself, a type of

carry over termed “stimulus generalisation” (Webster et al., 2015). There is some evidence to suggest this more desirable outcome was met, as pictures used in the outcome measure (ACT) were different from those used in Listen-In therapy. This demonstrates successful generalisation across different exemplars. In some cases, visual similarity could support this process (e.g. two different oranges will look visually similar); however, this is unlikely to account for the majority of improvements due to the large number of test items, and not all pictures were visually similar. Furthermore, although the treatment and ACT contained the same audio stimuli, speakers were randomly selected in each trial. In hindsight it would have been beneficial to include an outcome measure which assessed comprehension in, at a minimum, a different task, to asses carry-over of word comprehension across different contexts.

In summary, generalisation within aphasia treatments is a complex topic, and the factors which promote generalisation are still not well understood. The present findings are in line with word retrieval treatment studies which find that treatments targeting remediation of deficits do not tend to confer generalisable improvements to untrained items (Webster et al., 2015). The end goal of impairment-based treatments is to instigate change which benefits everyday communication. Although item specific effects have been criticised due to this lack of generalisation, as replicated here, one way of capitalising on item-specific effects is to carefully select treatment sets which will be relevant for individual patients. This was considered in the present study, where the treatment corpus was carefully curated to include highly frequent words in spoken English. A different approach would be to train smaller but more relevant treatment sets. In clinical settings curating individual word lists may be challenging, as patients can be unware of their deficits and which words they find difficult. One solution in Listen-In may be a mixed approach, where a pre-defined word list can be adapted to include personalised items.

5.2.3.1.6 Dose and treatment tolerance

Patients carried out a considerably larger dose of therapy in comparison to previous impairment-based treatment studies. Previous studies targeting speech comprehension have typically administered between 6-12 hours of therapy, and reported mixed and

inconclusive findings, in phonological based treatments (Archibald et al., 2009; Grayson et al., 1997; Hessler et al., 2010; Knollman-Porter et al., 2018; Maneta et al., 2001; Morris et al., 1996; Prins et al., 1989; Tessier et al., 2007; Z. V. Woodhead et al., 2017; Woolf et al., 2014), and also semantic based treatments (Bastiaanse et al., 1993; Behrmann & Lieberthal, 1989; Francis et al., 2001; Morris & Franklin, 2012; Munro & Siyambalapitiya, 2017). For example, Woolf and colleagues (2014) provided a total of 12 hours of treatment over six weeks to their eight patients, which was also divided amongst a number of different tasks, meaning time spent on any one task was low. Over the duration of the study, this equates to two hours of therapy per week, a low intensity treatment. In contrast, Listen-In patients achieved, on average, seven hours of therapy per week, corresponding to a high dose and moderately intensive treatment (~7 hours/week) (Stahl et al., 2018). One reason for positive treatment effects may be that this was a more sufficient dose, in line with recent reviews which demonstrate that high dose aphasia treatments tend to produce better outcomes (Bhogal et al., 2003; Brady et al., 2016). Bhogal and colleagues (2003) found that studies which delivered, on average, 98 hours of impairment-based intervention produced meaningful improvements in patients' language abilities. The most recent Cochrane review of Speech and Language therapy in post-stroke aphasia also concluded that there is a benefit of therapy when a dose of between 60-208 hours is delivered (Brady *et al*., 2016). The average dose achieved by Listen-In patients therefore falls within this previously identified range.

Previously, dose and intensity have not been well controlled or reported in treatment studies, therefore the differential effects of these two variables are unclear (Cherney et al., 2011; Dignam, Rodriguez, et al., 2016). For example, 7 hours per week, as achieved by Listen-In patients, may be considered both a moderate (Stahl et al., 2018) and low intensity (Dignam et al., 2015) treatment schedule depending on differing viewpoints. The present study did not set out to test a dose-response relationship, or compare intensity of treatment. It is important to note that dose was not manipulated as an experimental variable, as all patients were given the same target of 100 hours. Therefore it is possible that patient related factors may have influenced the propensity of patients for doing more or less training. However, these findings contribute to this debate by showing, for the first time, that speech comprehension treatment can be effective across a wide range of

patients when delivered at a high total dose, even years after stroke. The precise dose required for optimal Listen-In treatment is yet to be investigated.

Improvements were not monitored throughout 12-week treatment block, therefore it is unclear what the most optimal treatment dose or intensity may have been at a group and individual level. From the dose response scatterplot, it is clear that there is no simple linear relation between total dose and treatment outcomes; but this does not mean that a dose-response relation was not present. For example, it could be that dose is associated with treatment response within the first few sessions or hours, but then shows little relation thereafter. For example, Knollman-Porter and colleagues (2018) found rapid improvements over the first several sessions in their two aphasic patients during a similar speech comprehension treatment. Future studies may look at investigating optimal parameters of delivery for speech comprehension treatment, with the aim of pinpointing treatment schedules which promote the greatest change, but that are also achievable for patients. The present findings suggest that 100 hours over 12 weeks (or ~8 hours/week) is an achievable target for many patients in this type of computer-based treatment.

The aim of Chapter 1 was to develop a treatment app, with gamification, which was useable and enjoyable for patients with aphasia to use for prolonged periods of time. Many patients were able to achieve the target dose, suggesting this outcome was successful. However, a consistent comment by patients was that the therapy was repetitive, and that this negatively impacted on continued practice and enjoyment. These findings mirror those previously found (Varley et al., 2016), where patients achieved approximately 18 hours of treatment over 6 weeks on a self-administered app targeting apraxia of speech. Of note is that Listen-In patients achieved a considerably larger dose, and one reason for this could be the implementation of gamification to increase enjoyment. Six patients achieved considerably less than the target dose of 100 hours (<40 hours), despite frequent communication with the research team. Reasons included time constraints, change in medication and levels of fatigue and motivation, difficulties with self-initiating therapy at home, low motivation, and boredom (in a patient performing near to ceiling level). These factors were mainly patient related rather than issues with the task itself, therefore it is likely that, as in any therapy, some patients may not be well

suited to this type of treatment. However, for the majority of patients, high dose self-administered therapy was found to be achievable.

Another key finding was that gamification had a polarising effect; patients responded both positively and negatively to this component. It is therefore likely that other factors contributed to the overall high dose, such as contact with the research team, which included telephone calls and emails. In some cases, these were to motivate patients who had fallen behind in treatment, therefore the contribution of these communications should not be underestimated. Previous findings have suggested that intensive treatments entail higher drop-out rates, not found in the present study. One reason may be the flexible self-administered nature of treatment in contrast to scheduled face-to-face treatments. The patients enrolled in Listen-In may also form a particularly motivated group, as the majority had taken part previous research studies. For these reasons, out of the context of a research trial, it may be that patients do not achieve a comparable dose. However, by enabling patients to use the app independently, patients are put back in control of one aspect of their treatment, and are free to set their own goals regarding dose, without reliance on therapist time or resources. This also represents an economically viable option in light of the limited amount of time patients receive for speech and language treatment in the National Health Service in the UK (Code & Heron, 2003, 2009). Future work could look to develop the app, based on user feedback, to increase how enjoyable the experience is, and enable personalisation depending on individual preferences.

5.2.3.1.7 Limitations

An unexpected finding was that Group 2 were significantly better than Group 1 on the ACT at baseline, even though they had been randomised on the basis of CAT performance. Previous pilot testing, and baseline results in the current study, reveal a high level of concurrent validity between the ACT and CAT-words subtest, so this was surprising. The cause of this imbalance may be differences between the ACT and CAT-words. The ACT includes words presented in carrier sentences (versus on their own), has 5 foils (versus 3), has more word classes (verbs and adjectives as well as nouns) and scores only on accuracy

(versus accuracy and reaction times). Further to this, Group 2 outperform Group 1 on CAT-words performance at baseline (although not significant) indicating some degree of existing imbalance. In hindsight, with this knowledge, minimising on baseline ACT performance may have led to groups that were more equal. However, initial severity did not relate to treatment outcomes, and the study used a within subject design, therefore the impact of this difference appears minimal.

Another imbalance (unrelated to Group) is better performance on Set B versus Set A on the ACT (~6%). This is in line with results from the pilot testing phase. It is possible this is due to some items being inherently more difficult. However, a closer look at performance of healthy controls reveals that when repetition is taken into account in the scoring method, they perform significantly worse on Set A. This indicates the reason could be related to the audio files, or some degree of ambiguity in specific challenges, which requires repetition. This appears to be sufficient for healthy controls to choose a correct picture, but perhaps not PWA.

A limitation of the therapy program is that there was variability in the pattern of exposure of treatment stimuli over time, across the treatment block. Although Trained Challenge Items (the construction tested in the ACT) showed consistency over time, there was considerable variability in exposure of lexical items over time (trained items, but in a different linguistic construction to that tested in the ACT). Analysis of two patients illustrated this point by showing that some Trained Lexical Items were only exposed towards the beginning of treatment, and also that this variability differed across patients. One reason for this is that the algorithm was adaptive, so items with only 'easy' constructions were not returned to later in therapy for many patients. This may have impacted the results by biasing certain test items in two ways: by total exposure count, and time between exposures. Some patients reported that they got 'stuck' in repetitive cycles of challenges, therefore some may have been 'overexposed'. In the future, one possibility would be to control the exposure rate and frequency of items across the treatment block.

5.2.3.1.8 Future directions

Listen-In aimed to maximise dose over a period of 12-weeks, but there is evidence that less intense, or distributed therapies can also be effective (Cherney, Patterson, & Raymer, 2011; Dignam et al., 2015). A future important question, not answered here, is what is the most effective and efficient dose of Listen-In therapy is. If therapy can be delivered in a more flexible patient led manner, whilst adhering to scientific principles which maximise effectiveness, then these types of repetitive mass-practice tasks may become more tolerable. The trial version of Listen-In used a linear adaptive algorithm which moved patients 'up' and 'down' challenges according to difficulty level. Future work could look to build on this mechanism to track performance by lexical item and withhold or expose depending on past performance. In this way, therapy delivery may become more efficient and patients may spend less time on items they consistently perform well on.

A frequent report from patients and family and carers (including those who achieved high doses) was that the repetitive nature of therapy was tiresome. Although gamification was implemented to try to ameliorate this issue, it is likely that not all patients were motivated sufficiently to offset the repetitive nature of the spoken stimulus-to-picture matching task. One observation is that a target dose of 1 hour and 20 minutes each day may have been too high for some patients, and a lower dose may have been more tolerable.

Performance on treated items was not monitored over the 12 week period. It is possible that patients improved after the first several exposures, meaning subsequent exposures may not have been therapeutic. One future avenue would be to track item exposure and accuracy to ascertain the most effective exposure regimen of individual items.

5.2.3.1.9 Clinical implications

This study contributes to the limited evidence base by finding a significant group level improvement in speech comprehension following a computerised phonological-semantic therapy, in a relatively heterogeneous group of PWA. The automated and independent nature of this therapy program means it has the potential to be translated easily into

clinical practice. Listen-In will be released as a standalone app in the near future, therefore all patients with aphasia will be able to access this application. As there are currently no therapy apps which focus solely on speech comprehension, this will provide the first evidence based therapeutic app to target this level of impairment. As such it will make a significant contribution to clinical practice and to individual patients.

5.3 Chapter 3: Predicting response to Listen-In treatment using structural imaging

In Chapter 2, automatic linear modelling (ALM) showed that a combination of baseline factors explained some of the variability in treatment outcomes. However, this was small, and no one measure was strongly associated with treatment outcomes. The present chapter builds on this analysis by investigating whether structural integrity of brain tissue, prior to treatment, predicts Listen-In treatment outcomes.

Aim: To investigate whether structural integrity of pre-treatment grey matter (GM) and white matter (WM) predicts response to Listen-In treatment.

Hypotheses:

(1) Structural integrity of baseline GM and/or WM in the left hemisphere (LH) will correlate with change in spoken word performance from pre to post treatment.
(2) If significant correlations are found, these may be observed in the left hemisphere (LH) language network, in peri-lesional regions, where variability in integrity will be present amongst the patient group.

5.3.1 Experimental Procedures

The structural imaging data comprised of baseline MRI scans taken from T2 in the main Listen-In trial. The dependent variable used in these analyses was percentage improvement on treated items on the ACT from pre to post treatment. This was the significant behavioural finding from Chapter 2 which showed variability across the patient group.

5.3.1.1 Participants

Twenty-five patients from the main study cohort, who were able to be scanned, were included in the present analysis. Patients were representative of the main cohort in age, time since stroke, baseline speech comprehension performance (ACT) and response to

treatment (change on treated items) (Table 5-5). The subgroup of twenty-five patients were spread evenly across Group 1 (N=13) and Group 2 (N=12). Seven patients were scanned on a 1.5T scanner, and eighteen were scanned on a 3T scanner (see Methods).

Subgroup	N	Age	Time since stroke	Baseline ACT (T1)	Change treated items (%)
Behaviour only	10	56 (9)	48 (36)	57 (15)	9 (5)
Behaviour + MRI	25	62 (13)	87 (63)	52 (12)	11 (11)

Table 5-5 Demographic and behavioural variables for the subgroup of patients able to be scanned, and the subgroup of patients with behavioural data only. N=number of patients. Brackets are standard deviations of the mean.

Figure 5-17 displays the distribution of patient lesions across the whole brain. All patients had lesions involving the left temporal lobe. Maximal lesion overlap is observed in the temporo-parietal junction extending into middle and anterior temporal regions, inferior parietal regions, and frontal lobe. Twenty out of twenty-five patients had lesions in the region of Heschl's gyrus, and nineteen showed lesions consistent with patients with Wernicke's type aphasia, in the posterior temporal lobe in Brodman's area 22. Twenty-one patients showed lesions which extended to the inferior frontal gyrus (IFG), including Brodman's area 45, consistent with damage to Broca's area. Twenty had lesions to medial subcortical structures. In summary, the majority of patients had large and extensive lesions which encompassed extensive parts of the LH. Three patients had additional RH lesions, which occurred before their aphasic LH stroke, in the following regions: fronto-parietal and occipital lobe (P24), fronto-parietal lobe (P25), and lesions consistent with anterior and posterior middle cerebral artery stroke, and left cerebellum (P32).

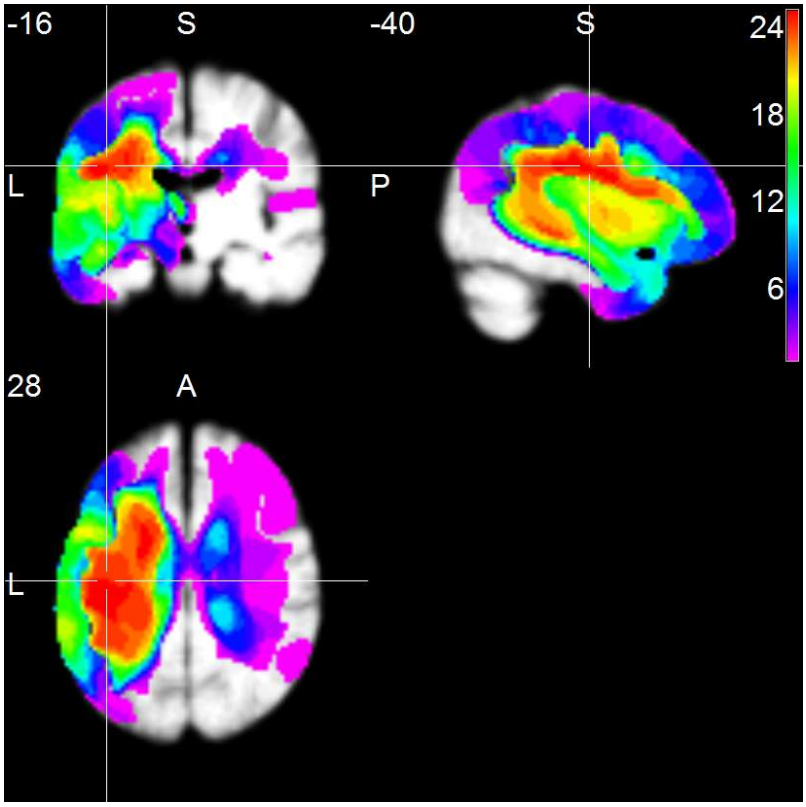


Figure 5-17 Lesion overlap map showing distribution of lesions across the whole brain. The colour bar represents the number of patients with a lesion in a given voxel, from 1 (pink) to 25 (red). Numbers represent MNI coordinates.

5.3.1.2 Data preprocessing

Images were smoothed with an isotropic kernel of 10mm full-width half-maximum. For details of image acquisition and prior preprocessing steps see Method.

5.3.1.1 Statistical analyses

Twenty-five smoothed and modulated GM and WM images were then entered into two separate multiple linear regression models in SPM12. These analyses was designed to identify regions where volume of GM and WM correlated with Listen-In treatment response. Percentage improvement on trained items on the ACT, from pre to post therapy, was used as the dependent variable, as this was the significant behavioural effect from Chapter 2. Effects of age, time since stroke and lesion volume were modelled by including them as covariates of no interest. In keeping with standard practice, the statistical voxel-level threshold was set at $p< 0.001$ with cluster-level significance set at

$p < 0.05$ after family-wise error (FWE) correction for multiple comparisons across the whole search volume (using random field theory as implemented in SPM; Flandin & Friston, 2015).

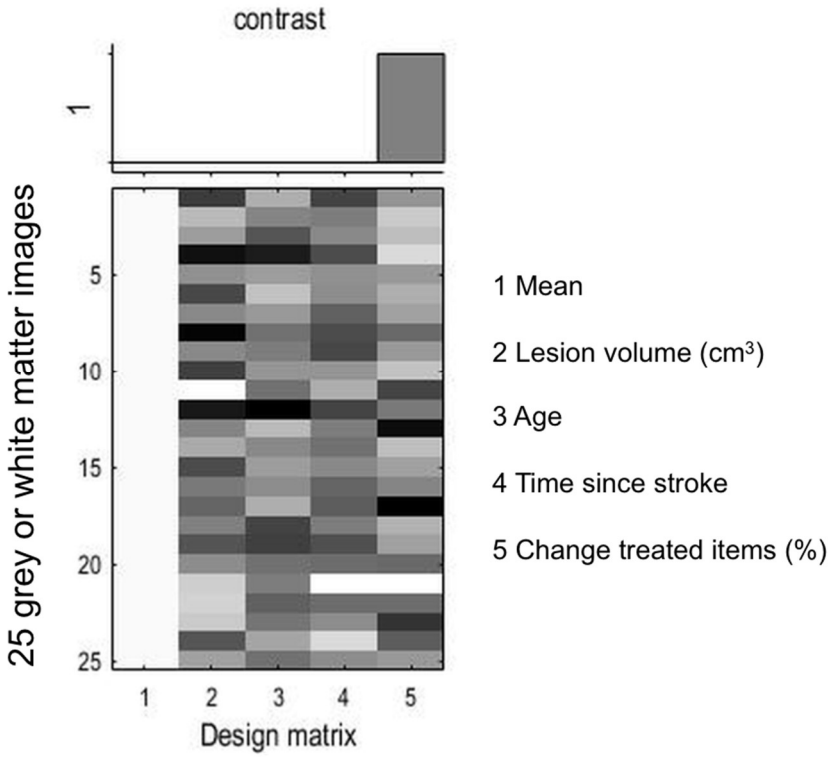


Figure 5-18 Design matrix used in GM and WM multiple linear regression models in SPM12. Both models included one regressor of interest: (5) change in treated item performance from pre to post therapy; and four regressors of no interest: (2) lesion volume (cm^3); (3) age; (4) time since stroke.

5.3.2 Results

5.3.2.1 Voxel based morphometry

Figure 5-19 displays results of the VBM analysis, demonstrating where greater volume significantly covaried with greater treatment response across the patient group. Results are displayed at the cluster level. The results show one GM cluster, and four WM clusters, where greater volume at baseline positively correlated with greater improvements in speech comprehension following Listen-In treatment. The locations of these clusters are:

- (1) GM located in the right basal ganglia, at the head of the caudate nucleus, with some overlap with the globus pallidum.
- (2) WM intrinsic to the right ventrolateral prefrontal cortex (VLPFC) underlying the IFG. Also, WM overlapping with the genu of the corpus callosum (CC) and anterior corona radiata in the right. Also, WM underlying the anterior cingulate cortex (ACC) bilaterally.
- (3) WM intrinsic to the anterior temporal lobe (ATL), underlying the middle temporal gyrus (MTG), anterior inferior temporal gyrus (ITG), and middle superior temporal gyrus (STG).
- (4) Large WM cluster connecting the right fusiform cortex with the thalamus, overlapping with regions of the hippocampus and parahippocampus, and inferior longitudinal fasciculus (ILF)/inferior fronto-occipital fasciculous (IFOF).
- (5) Bilateral WM intrinsic to the cerebellum, mainly in the LH.

	Cluster size (k_E)	t-value	Peak MNI coordinates (mm)			Brain structure	Brodmann's area
			x	*y*	*z*		
Cluster 1 GM	435	4.94	12	12	0	R caudate (head) / globus pallidum	R: BA25
Cluster 2 WM	2415	5.39	44	32	6	*Highest peak:* R ventrolateral prefrontal cortex	R: BA45, BA47
		5.06	14	36	2	*Second highest peak:* R: Genu of CC, anterior corona radiata L: WM underlying ACC	
Cluster 3 WM	1589	4.11	64	-12	-26	R MTG / anterior STG / anterior ITG	R: BA20, BA21, BA48
Cluster 4 WM	2050	4.90	30	-38	-8	R fusiform gyrus / parahippocampus / hippocampus / thalamus / caudate / ILF / IFOF	R: BA20, BA27, BA37
Cluster 5 WM	822	5.54	-6	-62	-28	Bilateral cerebellum	R: BA37

Table 5-6 Significant clusters from cross-sectional voxel-based morphometry analysis. Voxel-wise threshold set at p<.001, FWE correction at p<.05. Coordinates are the peak voxel with the highest t-score in that cluster. GM=grey matter, WM=white matter, R=right, L=left.

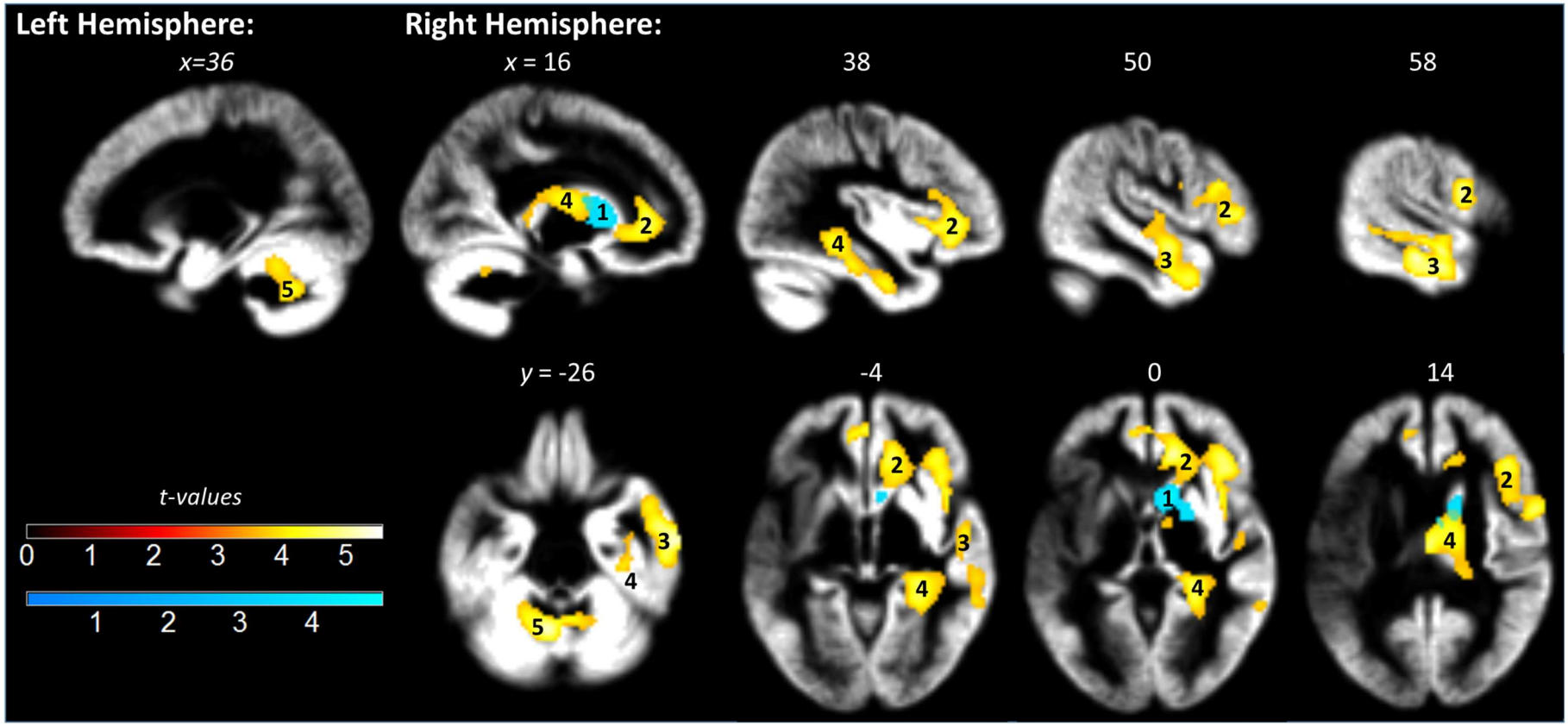


Figure 5-19. Standard T1-weighted anatomical slices showing regions of GM and WM where volume significantly covaried with change on trained items (%) from pre to post therapy. GM is in blue, WM is in yellow. Clusters are overlaid on an average GM template from 25 patients. Voxel-wise threshold set at p<.001. All clusters survived FWE correction at p<0.05. The numbered regions are as follows: (1) right caudate nucleus; (2) Right: WM intrinsic to VLPFC, genu of CC, anterior corona radiate; Bilateral: ACC; (3) WM intrinsic to anterior temporal lobe; (4) WM connecting the fusiform gyrus with the thalamus, overlapping the hippocampus, parahippocampus, ILF, IFOF; (5) WM intrinsic to bilateral cerebellum.

5.3.2.2 Tissue volume and treatment response

In the previous analyses, greater brain matter volume in mainly the right hemisphere was found to covary with greater response to therapy, across five regions. Three patients within the sample had additional right hemisphere lesions. One explanation for this findings may be that these patients are driving these findings; in other words, patients with less volume in the RH due to lesion damage may respond less well to therapy, accounting for the correlation. To examine this question, two correlations were conducted, one with the full sample of 25 patients, and one with only the 22 patients without additional RH lesions. Correlations investigated whether total volume of GM and WM, in the significant regions previously identified (five clusters), covaried with response to treatment.

Total volume of these five regions were first extracted from each patient's baseline segmented and modulated GM and WM images. To extract volumes of these significant regions, cluster results from the previous analyses were converted into binary masks. The masks were then used to extract volume from patients' baseline segmented and modulated GM and WM images, only within these regions of interest. This produced one GM cluster volume, and four WM cluster volumes, for each patient. Volume in each of these five regions was totalled separately for GM and WM, to form a single GM and WM volume (millilitres^3) for each patient. These values were then entered into four group wise correlations, where GM and WM were correlated with patients' improvement on trained items (N=25, N=22). A more stringent alpha value was set at $p \leq .013$ based on a Bonferroni correction for 4 comparisons ($\alpha = .05/4$).

As expected, results show that GM and WM volume highly correlate with treatment response for the full sample of 25 patients, in line with the previous VBM findings (Table 5-7). To investigate whether three patients with additional RH lesions were driving this result, the correlation was repeated with these three patients removed (N=22). Results show that the correlation was maintained. However, the correlation between GM and treatment response was no longer significant with these three patients removed.

Scatterplots in Figures 5-20 and 5-21 display the relation between volume of GM and WM clusters, and treatment response.

	Cluster volume	
	GM	WM
Change on treated items (%) N=25	.53*	.68**
Change on treated items (%) N=22	.50	.68**

*Table 5-7 Spearman's correlations between GM and WM cluster volume and change in treated item performance (%) from pre to post therapy. N=25 is the full cohort, N=22 is a subgroup with three patients with additional right hemisphere lesions excluded. Alpha value of $p \leq .013$ set following Bonferroni correction for multiple comparisons ($\alpha = .05/4$). *Significant at p<.01, **significant at $p \leq .001$.*

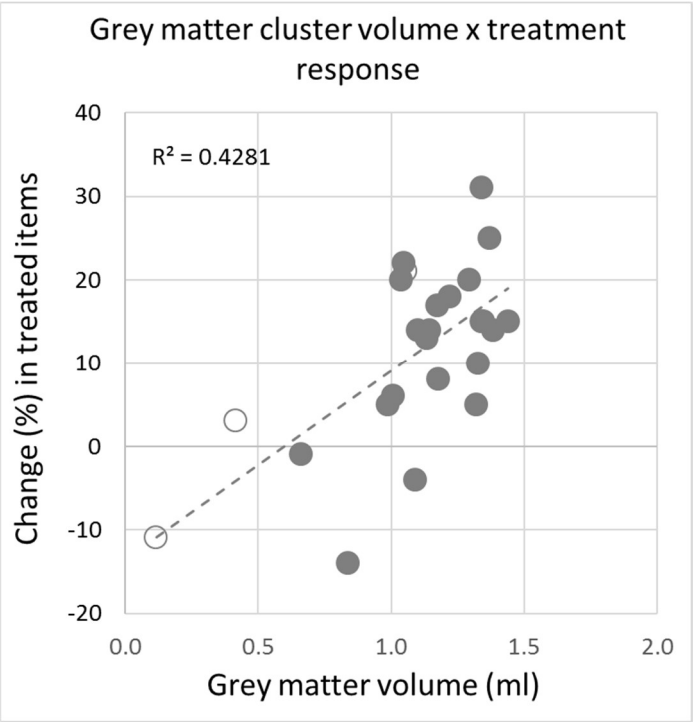


Figure 5-20 Scatterplot showing the relation between grey matter cluster volume and treatment response for 25 patients. Unfilled circles are patients with additional RH lesions.

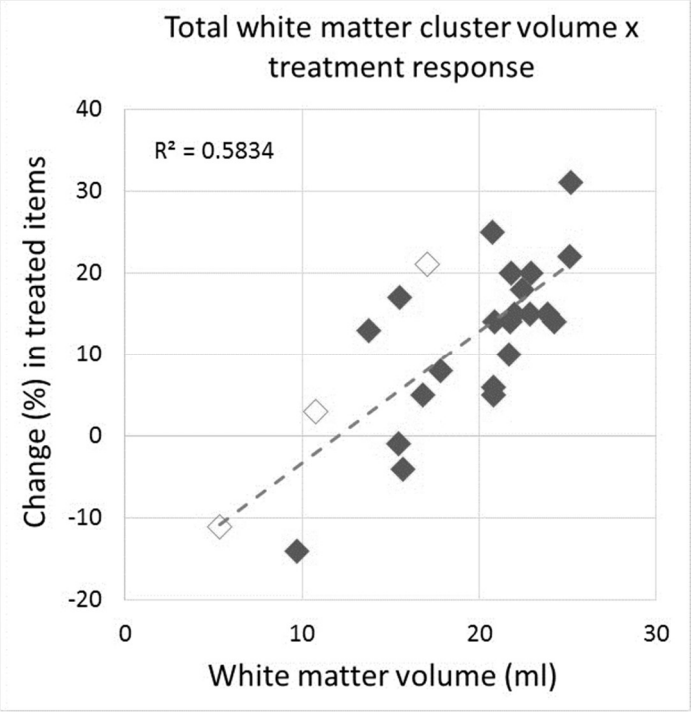


Figure 5-21 Scatterplot showing the relation between total white matter cluster volume and treatment response for 25 patients. Unfilled diamonds are patients with additional RH lesions.

5.3.2.3 Caudate nucleus overlap map

Cluster 1 was located in the head of the caudate nucleus, a region with extensive connectivity with the cortex. To examine this further, this cluster was overlaid with cortical connectivity maps derived from diffusion tractography with healthy participants, displayed in Figure 5-22 (Draganski et al., 2008). The cluster identified in the present study demonstrates the greatest overlap with regions of the caudate which has connectivity with the dorsolateral prefrontal cortex (DLPFC).

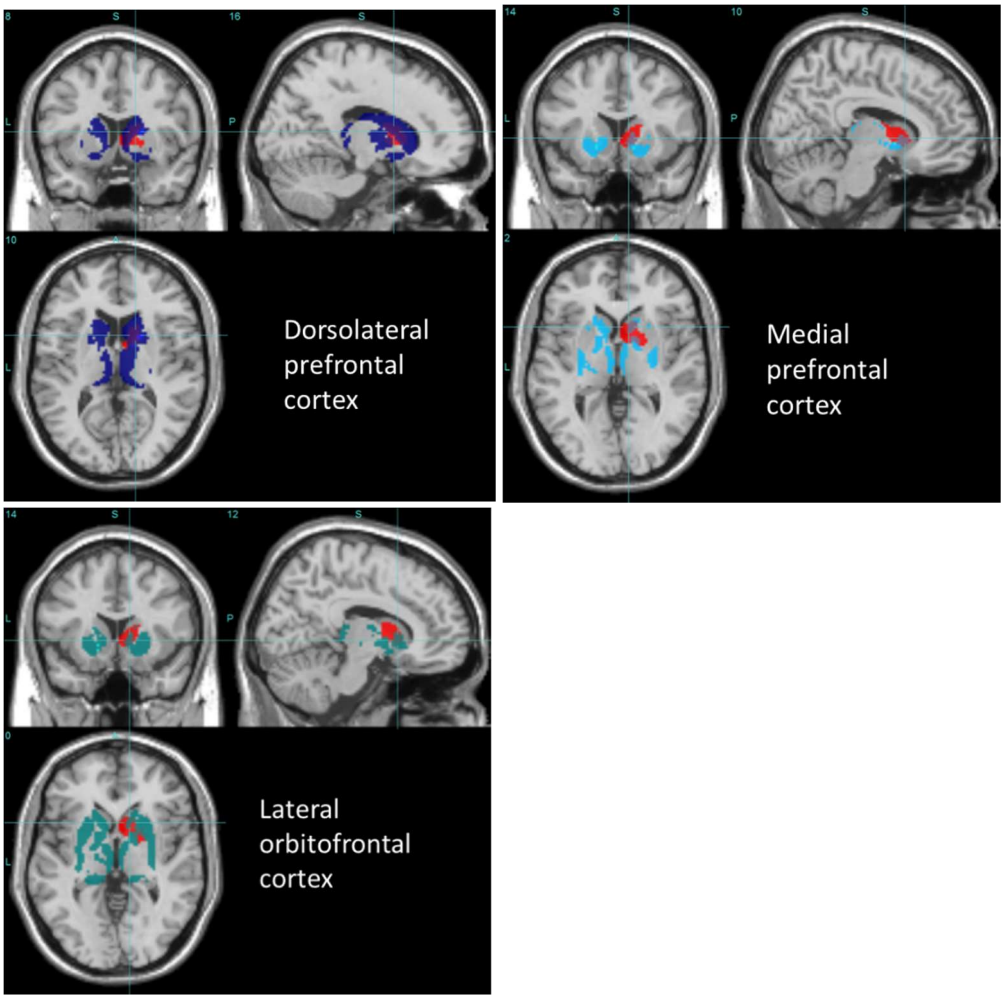


Figure 5-22 Overlap map showing regions where the right caudate nucleus cluster (red) overlaps with cortical connectivity maps of the basal ganglia (Draganski et al., 2008). Maps are overlaid on a single subject template image in MRIcron.

5.3.3 Discussion

The purpose of this study was to examine whether integrity of pre-therapy brain structure could predict Listen-In treatment outcomes in patients with chronic aphasia. After controlling for age, time since stroke, and lesion volume, the results revealed one GM cluster, and four WM clusters, predominantly in the RH, where tissue was positively associated with therapy outcomes; patients who entered Listen-In treatment with greater volume of these regions prior to therapy, tended to make larger gains on treated items post-therapy. These findings support Hypothesis 1.

In Hypothesis 2 it was speculated that significant regions would be found in peri-lesional regions in the LH language network. This hypothesis was not supported. Increased volume was found to relate to better treatment outcomes, but significant regions were located predominantly in the RH. Two clusters were present in the LH: one cluster was in the cerebellum, suggesting connectivity with RH regions; and the second cluster was in WM underlying the ACC, anatomically remote from classic language regions.

Most lesion-symptom mapping approaches have identified regions in the LH where damage is associated with speech comprehension deficits (Bates et al., 2003; Dronkers et al., 2004). Furthermore, almost all previous studies have focused their investigation on regions of damage in the LH. The design of the current study differed as it related structural integrity pre-therapy, to change in performance post-therapy, rather than to speech comprehension ability per se, and carried out a whole brain analysis. In this way, it is suggested that the current findings demonstrate regions which appear to be important for both linguistic and cognitive functions, which subserve learning during this type of task.

5.3.3.1 Comparison with previous findings

Left hemisphere brain structure and treatment outcomes

A handful of previous studies have reported on the role of pre-therapy brain structure on treatment outcomes (Aguilar et al., 2018; Fridriksson, 2010; Marcotte et al., 2012; Meinzer et al., 2010). Four of these studies reported investigations in lesioned regions in the left-hemisphere. Using a lesion-symptom mapping approach, Fridriksson (2010) found that patients with damage to the left middle temporal lobe were less likely to improve following anomia treatment. Marcotte and colleagues (2012) used a correlational analysis and found that damage to Broca's area was strongly negatively associated with naming improvement, in that those who had more damage to this region tended to make the least improvements. In a reading treatment study, Aguilar and colleagues (2018) partitioned patients' lesion data into anatomical regions of interests, and found that WM in the LH in Broca's area, the insular, inferior longitudinal fasciculous (ILF), and WM connecting the thalamus to the parietal cortex, all contributed to variability in response to a computerised reading treatment (alongside behavioural measures). The latter findings show similarity with the present results as they indicate that integrity of WM pathways influence treatment response. However, in Aguilar and colleagues (2018) patients these were LH WM pathways which were likely to have been directly disrupted by the lesion itself, whereas the present findings were found in the RH. Although anatomically remote from the lesion site, it is possible that lesion effects extend to these regions through secondary processes such as Wallerian degeneration, therefore although not directly affected by the lesion, the effects of the lesion on these regions cannot be easily separated out. In further a novel study, integrity of the hippocampus in the LH was found to relate to treatment outcomes (Meinzer et al., 2010). Similarly, the present study found WM regions in the vicinity of the hippocampal area were related to response to therapy, but in the RH.

These studies focused their analyses on lesion data, or particular LH structures, and did not report investigations into the residual cortex, so it is not clear what role RH brain structure may have had in their patients' outcomes. In contrast to these findings, the current study failed to find LH regions where integrity covaried with treatment outcomes. It is possible that the VBM analysis was insensitive to peri-lesional areas due to variability in lesion site amongst patients. The lesion overlap map demonstrates patients had large and overlapping lesions, and in this way, there may have been insufficient residual cortex

to identify regions which contributed to treatment outcomes. Due to the exploratory nature of the present findings, an unbiased whole brain approach was selected; however, future analyses using regions of interest, such as parcellation techniques (Aguilar et al., 2018), may yield further insight into the role of LH regions in Listen-In treatment outcomes.

Right hemisphere white matter volume and treatment outcomes

The scatterplot shows that the association between GM and WM volume and treatment response was positive, and the subgroup correlation showed this was not driven solely by patients with additional RH lesions. As the association was linear, it was not the case that WM volume categorised patients into responders or non-responders. This suggests GM and WM volume is indicative of some kind of aptitude for treatment, or capacity for learning, which varies amongst participants.

Previously, cross-sectional studies which have investigated WM in healthy individuals have explored differences in WM volume (or integrity) between groups, or correlated WM structure with behavioural measures. These studies have tended to identify localised regions of WM which underlie the particular skill being investigated. For example, using VBM, literate adults were found to have greater WM volume in the splenium of the corpus callosum, part of the reading network, compared to non-literate adults (Carreiras et al., 2009); and in a study utilising diffusion tensor imaging (DTI), an association was found between greater learning ability in an artificial grammar task, and greater WM integrity in Broca's area (Flöel, de Vries, Scholz, Breitenstein, & Johansen-Berg, 2009). In a further training study, Golestani, Paus, and Zatorre (2002) also used VBM to investigate response to a training task in a group of healthy individuals learning non-native phonetic contrasts in Hindi, and identified a bilateral region in the parieto-occipital sulcus where greater WM volume correlated with faster learning rate (in addition to morphological differences). In line with the studies described, the present findings demonstrate an association between greater volume of WM, and greater behavioural performance, supporting the assumption that greater volume is associated with greater functionality, and vice versa. However, caution is needed when comparing findings from healthy individuals and stroke patients,

due to important differences in brain structure and function between these two groups; it is not possible to directly compare small differences in WM structure in healthy individuals, with the extent of WM loss amongst stroke patients. However, findings from healthy individuals, although not directly comparable, are important to consider, as they provide findings which link brain structure to behavioural performance.

In contrast to previous studies which have identified localised regions, the present study identified distributed regions throughout mainly the RH, in both cortical and subcortical structures, which predicted treatment outcomes. The role of the RH in aphasia recovery is unclear, but a typical finding in functional imaging studies is increased task-based activation in RH areas homotopic to LH language networks (Turkeltaub, Messing, Norise, & Hamilton, 2011), suggesting functional compensation, or reorganisation, within the RH. However, whether RH activation reflects adaptive or maladaptive mechanisms for language recovery is controversial (Cocquyt et al., 2017). For example, activity in the RH, pre-therapy, has been found to predict treatment outcomes (Richter et al., 2008); but strong right lateralisation has also been associated with language declines (Breier et al., 2009). It is likely that a number of factors interact during recovery making individual recovery trajectories heterogeneous; so far, group studies may be insensitive to these differences (Abel et al., 2015). Nevertheless, the present findings offer a novel contribution to this debate by demonstrating that greater WM volume in the RH facilitates outcomes during this type of treatment.

Few studies have specifically investigated how brain structure in the RH may contribute to recovery. One study which did investigate this found increased GM volume, compared to healthy controls, in localised regions, in the supplementary motor area (SMA) and MTG, that were related to language production and comprehension respectively, suggesting neural adaptation in the RH post-stroke (Lukic et al., 2017). The present study builds on these findings by showing that WM volume, as well as GM, is also indicative of later treatment outcomes. Unlike the previous study, RH WM structure was associated with treatment outcomes, rather than language ability. These measures entail key differences: Lukic and colleagues (2017) conclude that the RH supports language functions following LH lesion damage, as their measures were linguistic ability; instead, the present findings

are in line with those previously discussed which link WM structure with learning ability rather than language skill. For example, in Golestani and colleagues (2002) participants', WM was related to learning rate but not skill level, and initial skill level did not relate to learning rate; in the present study, speech comprehension ability also bore no relation to treatment outcomes. These findings suggest that the RH is supporting treatment in functions beyond language specific processes.

However, a different interpretation of Lukic and colleagues (2017) findings, which the authors acknowledge, is pre-morbid individual differences. It may be that variations in brain structure, prior to stroke, exert an influence over aphasia symptoms, as well as language recovery. For example, it is possible that greater lateralisation of white matter pathways and language functions in the right hemisphere, may provide an advantage for some patients, if the right hemisphere is better able to support or take over following left hemisphere damage. In Lukic and colleagues (2017) patients is not possible to differentiate between greater volume due to post-stroke neuroplasticity, or greater volume due to pre-stroke individual differences. In the same way, the present finding may also be accounted for by both of these interpretations. Nonetheless, the correlation between right hemisphere volume and response to treatment suggests that the right hemisphere plays a role in re-learning during treatment, regardless of the underlying aetiology of volume differences amongst individuals.

Relatively fewer studies using VBM have reported WM findings. This may be because investigations directed at WM typically utilise diffusion based imaging techniques (e.g. Flöel, de Vries, Scholz, Breitenstein, & Johansen-Berg, 2009; San Chen, Han, Qi, Gabrieli, & Garel, 2014), which are more sensitive measures of white matter integrity (Le Bihan, 2006). However, several previous studies have shown that VBM can be sensitive to differences in white matter volume in other populations, including individuals who stutter (Jäncke, Hänggi, & Steinmetz, 2004) individuals with dyslexia (Dole, Meunier, & Hoen, 2013), and persons with Alzheimer's disease (Hugenschmidt et al., 2008). The present results build on these findings by showing that VBM, with T1 weighted images, can be sensitive to white matter variations in stroke patients.

5.3.3.2 Distribution of clusters within the right hemisphere

Two clusters correspond to classic language regions in RH homologues, in lateral frontal and temporal cortices. In healthy individuals, these regions respond to linguistic input across modalities, and have been shown to be robust across and within subjects over time (Fedorenko, 2014). The first cluster comprised WM in the frontal lobe which had two main peaks. These were in WM underlying the ventrolateral prefrontal cortex, and WM overlapping with the genu of the corpus callosum and anterior corona radiata.

The lateral portion of this cluster underlies the IFG, a homologue to Broca's area which has traditionally been linked to language production. The IFG has distinct morphological and functional regions. The left IFG has been implicated in semantic and phonological tasks and working memory, in tasks such as associative learning and verbal memory, skills likely to be engaged during Listen-In (Liakakis, Nickel, & Seitz, 2011; Mestres-Missé, Càmara, Rodríguez-Fornells, Rotte, & Münte, 2008). Notably, these functions are left lateralised (Liakakis et al., 2011). However, aphasic patients have been shown to reliably recruit the right IFG across multiple language tasks, in contrast to LH networks in controls (Turkeltaub et al., 2011), pointing to a compensatory role of this region in language processing. The medial portion of this cluster overlapped with the genu of the CC, a rich bundle of nerve fibres which link the two hemispheres, and the anterior location suggests this region is involved in information exchange between the frontal lobes. The function of the anterior portion of the CC is unclear in relation to language, and is not typically identified within classic language models. However, some studies suggest a link with RH language functions. Schlegel and colleagues (2012) found changes in the anterior genu of the CC, bilaterally, in foreign language learners, suggesting these nerve fibres play a role in transfer of language information between the frontal lobes. The anterior portion has also been linked to prosody during speech comprehension (Friederici, von Cramon, & Kotz, 2007).

The second cluster was located in WM underlying the right ATL, including anterior and middle STG, anterior ITG, and anterior MTG. These regions correspond to anterior

portions of the ventral 'what' pathway in speech comprehension in the RH, although the extent of bilateral processing is currently debated (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009). The ATLs have been proposed to be a transmodal representational "hub" for semantic processing, with the left being specialised for verbal information, and the right for visual information (Rice, Caswell, Moore, Hoffman, & Lambon Ralph, 2018), and in aphasic patients, semantic performance has been found to relate to integrity of the left ATL (Butler et al., 2014). Furthermore, greater volume of GM in the right MTG (and insular) was found to correlate with spoken word comprehension in a group of 40 aphasic patients, suggesting the MTG in the RH supports word comprehension (Lukic et al., 2017). In a further study with aphasic patients, Hope and colleagues (2017) found structural adaptation in the ATL in the RH over the period of one year or more, which was related to improved language skills, suggesting this region may show adaptive compensation.

Taken together, these clusters correspond to RH regions which have been implicated in healthy speech processing, and speech processing in aphasic patients. One possibility is that these WM regions represent RH language homologues, supporting treatment in a compensatory manner. For example, Richter and colleagues (2008) previously found that activation of typical language networks in the RH was associated with treatment outcomes in their chronic aphasic patients, and the relation was linear, demonstrating individual variability in the functioning of these networks across patients, prior to treatment. In the same way, variability in WM volume underlying language related regions here may represent variability in functioning of typical RH language networks, important for treatment success. Importantly, in Richter's (2008) study, both patients and controls showed activation of the same RH regions, leading to the suggestion that typical RH regions were recruited by aphasic patients. A second not mutually exclusive possibility is that WM variability reflects differences in post-stroke adaptation. This would be in line with studies which suggest facilitatory structural adaptation in the RH in chronic aphasic patients (Hope et al., 2017; Xing et al., 2016). These possibilities will be returned to in more detail later. Either way, one suggestion is that some patients have better functioning language networks in the RH to support language processes, which could then facilitate treatment. This is supported by the present findings due to the location of WM clusters which underlie classic language regions in the frontal and temporal lobes.

Cluster 2, in the frontal lobe, overlaps in the right with the VLPFC, genu of the CC, anterior corona radiate, and in the left, underlies the ACC. The ACC is part of the cognitive control network, and has been implicated in a number of executive tasks including selective attention, working memory, and language generation (Abutalebi & Green, 2007). The VLPFC has also been implicated in executive functions to do with language, such as semantic working memory (Wagner et al., 2002). Cluster 4 comprised deep WM connecting the anterior fusiform gyrus with the thalamus, overlapping with hippocampus, parahippocampus, ILF, and IFOF. Taken together, these clusters correspond to memory networks involved in encoding and consolidation (Bubb, Kinnavane, & Aggleton, 2017), and more specifically, in word learning (Shtyrov, 2012). Previously, the ILF and IFOF have shown structural connectivity to key regions in bilateral speech comprehension networks (Turken & Dronkers, 2011). The single GM cluster (Cluster 1) was located in the head of the caudate nucleus in the basal ganglia, a region with connectivity to nearly all parts of the cortex (Grahn, Parkinson, & Owen, 2009). Overlap with connectivity maps showed that this part of the caudate has connectivity to the DLPFC, forming a cognitive cortico-striatal loop (Draganski et al., 2008), consistent with its role in executive tasks. In language, it is thought to be involved in selection and inhibition (Robles, Gatignol, Capelle, Mitchell, & Duffau, 2005), and there is evidence to suggest that damage to the left caudate nucleus contributes to aphasic impairments in stroke patients (Grönholm, Roll, Horne, Sundgren, & Lindgren, 2016), indicating a role in language functions. The final cluster (Cluster 5) was in bilateral WM in the cerebellum, with a strong leftward asymmetry. As all other clusters were located in the RH, this is consistent with crossed cerebello-cortical connectivity, forming part of WM networks in the RH. The cluster overlapped with the body of the cerebellum (rather than particular lobules) and so a function cannot be inferred; however, the cerebellum in general has been implicated in a range of motor, cognitive and language tasks (Stoodley & Schmahmann, 2009), including language production and perception (Callan, Kawato, Parsons, & Turner, 2007), although these are generally left lateralised. Interestingly, GM volume near to this cluster has been positively associated with spontaneous speech outcomes in aphasic patients in addition to the pSTG in the RH, suggesting the left cerebellum may play a role in language recovery, as part of RH networks (Xing et al., 2016).

It is clear from the range of regions identified that the present findings are likely to represent multifunctional networks, which underlie a plethora of linguistic and non-linguistic cognitive functions. Given the structural nature of these findings, it is not possible to assign particular functions to these clusters. Nevertheless, in the following sections, I will discuss how these may relate to networks which are likely to be recruited during the Listen-In task.

5.3.3.3 Learning new words: comparison with word learning networks

All patients had impairments in single word comprehension, therefore Listen-In required patients to encode, learn, and consolidate unfamiliar, or partially familiar, words, in order for these words to become familiar items stored within the mental lexicon. One way of conceptualising this learning process is through models of word learning in healthy individuals (Shtyrov, 2012). In this framework, a central control system has been proposed, composed of domain general processes (such as attention), which regulate local associative learning processes (Chein & Schneider, 2005). In Listen-In, patients needed to draw on domain general processes, such as selective attention, working memory, and performance monitoring, whilst also making associative links between phonological representations and their meaning, a domain-specific function.

Word learning networks are different from classical language networks. Rapid learning of new words has been attributed to activity in the left hippocampus, whilst later consolidation has been attributed to the hippocampus, neocortex, and subcortical structures (Shtyrov, 2012). In healthy individuals, word learning tasks have activated mainly left-hemisphere networks. These regions have included inferior, middle and medial frontal cortices and subcortical areas (Mueller et al., 2014), hippocampus, fusiform gyrus and inferior parietal lobe (IPL) (Breitenstein et al., 2005), IFG, middle frontal gyri, MTG, STG, parahippocampus, pre-SMA, thalamus, and bilateral caudate nuclei (Mestres-Missé et al., 2008). These domain general networks are thought to mediate the initial stages of learning; as practice increases, these activations then decrease (Mueller et al., 2014). Of note is that atrophy of WM in bilateral hippocampi and adjacent temporal

cortex is characteristic of patients with Alzheimer's disease, who have a primary deficit in encoding and consolidation, demonstrating a link between structural integrity of these regions and learning ability (Li, Pan, Huang, & Shang, 2012).

In a novel study, Pohl and colleagues (2017) emulated aphasia therapy in healthy individuals, and examined brain activation patterns before, during, and after training. In the computer-based therapy program, participants were required to learn associations between pseudowords and objects, similar to the Listen-In. Functional MRI results during the learning phase versus consolidation phase showed activation of bilateral caudate and right posterior MTG, whilst the consolidation phase showed increased activation in sensory-motor areas, left insula, and right STG. Furthermore, there was increased fronto-temporal activation for items which were successfully learnt, versus items which were not. The authors suggest that the caudate is involved in the learning process through executive control of language (in response selection and sequencing) and that the STG may be involved in consolidation by integrating pseudowords into the lexical network. Of note is that these findings show RH recruitment, versus previous studies which have identified mainly left lateralised networks, suggesting activation patterns may relate to specifics of the task being used. Some caution is therefore required in generalising these findings to broader word learning models. However, the implication is that the RH is likely to be involved in word learning in some respect, and therefore the findings in the present study are in line with this assumption.

Taken together, these previous findings demonstrate recruitment of domain general and domain specific networks during word learning, and the relative contributions of these networks are likely to be dynamic throughout the learning process. In the present study, there is a striking correspondence of clusters to these regions. It may be that integrity of WM which supports language processing through domain general functions (e.g. memory network, caudate) is important for the initial stages of learning, whilst fronto-temporal language networks are likely to be important for integration of unfamiliar words into the lexical network. In sum, the evidence demonstrates that a wide network of brain regions appear to be important for successful word learning. In Listen-In patients, greater volume may indicate greater functioning of these networks, leading to greater treatment success.

This interpretation may account for the widely distributed findings observed here, in subcortical white matter as well as fronto-temporal language regions.

Of note is the left dominance of word learning networks, in line with left dominance for language functions. Accordingly, hippocampal activity has been found to dissociate between verbal memory in the left, and visual memory in the right (Kelley et al., 1998). In line with this, Meinzer and colleagues (2010) demonstrate a link between structural integrity of the left hippocampus and surrounding WM, and outcomes following intensive anomia treatment. However, there is evidence that patients may also engage RH hippocampal networks during language tasks. For example, temporal lobe epilepsy patients (with left hippocampus damage) have been shown to recruit the right hippocampus in verbal memory tasks, suggesting hippocampal structures have a high degree of plasticity (Richardson, Strange, Duncan, & Dolan, 2003). In aphasic patients, there is also evidence to suggest involvement of the right hippocampus during an associative learning naming treatment (Menke et al., 2009).

5.3.3.4 Cognitive functions during speech comprehension

A further way these WM regions may support treatment during Listen-In is through cognitive control of language processes, which relate to the selection and temporal processing of language representations (Abutalebi & Green, 2007; Fedorenko, 2014). It is likely these processes overlap with those engaged during word learning, as previously described. In Abutalebi and Green's (2007) model, these LH cognitive control regions comprise the prefrontal cortex, ACC, basal ganglia, and IPL; the former three regions were all identified as clusters in the present analysis, but in RH homologues. Previously, activations in these regions have been reported during language tasks in aphasia patients. For example, in an fMRI treatment study, Marcotte and colleagues (2018) found recruitment of the right ACC and caudate nucleus in their two aphasic patients respectively, which the authors speculate relate to error suppression in naming. In a further example, Martin and colleagues (2017) delivered constraint induced aphasia therapy (CIAT) to 11 aphasic patients, and used fMRI to investigate changes in activation

patterns. Improvements in naming from pre to post treatment were found to relate to activity in the RH, in the ACC, cingulate, thalamus, cerebellum and temporal gyrus. The authors speculate that the ACC may play a critical role in recovery by "regulating" information to compensate for damage to typical language functions. These findings, and others, suggest that activity in RH networks support language functions through cognitive control mechanisms.

It has previously been proposed that aphasic patients may engage cognitive control networks, or cognitive strategies, to a greater extent during speech comprehension due to the effortful nature of language processing (Brownsett et al., 2014; Meltzer, Wagage, Ryder, Solomon, & Braun, 2013). For example, Meltzer and colleagues (2013) investigated response to a sentence comprehension task using MEG in their aphasic patients. They found comprehension correlated with activity in right posterior temporo-parietal regions, but during a delay period, activity extended to parieto-frontal areas, not seen in controls. The authors speculate that this relates to "effortful reprocessing" of language input in short term memory. Brownsett and colleagues (2014) found a similar result in their fMRI investigation of the salience network in aphasic patients and controls. All patients in Listen-In had impaired speech comprehension, making the task effortful; it is therefore plausible that increased reliance on cognitive control networks to support language processing may play a role in response to treatment. In this way, variation in WM volume observed here may reflect, in part, variability in functioning of these networks amongst patients. In other words, patients with greater volume of these networks may be better able to engage with and learn from treatment using these cognitive functions.

5.3.3.5 Comparison with functional imaging treatment studies in aphasia

A number of studies have investigated functional networks recruited during language treatment in aphasic patients, and how these relate to treatment outcomes. Although these are functional imaging studies, the correspondence of the present structural findings to WM regions strongly suggests that treatment outcomes are related to the

functioning of WM networks. Two of these studies, illustrated below, provide insight into why these RH WM regions may have covaried with treatment outcomes in the present study.

In a training study similar to Listen-In, Menke and colleagues (2009) delivered an intensive computerised naming treatment to eight aphasic patients over two weeks, using an associative learning paradigm with auditory and visual cues. Using fMRI the authors found that short term success was related to activity in bilateral hippocampal formation and fusiform gyri, and right precuneus and cingulate gyrus, which they attribute to memory encoding, attention and multimodal integration. Long term success was related to activity in Wernicke's area, and to a lesser degree in perilesional temporal areas in the LH. The authors suggest that RH regions may be recruited due to effortful processing, with right hippocampal regions mediating recruitment of RH language homologues. The present findings offer evidence which supports this account. Pre-therapy volume of WM, underlying the hippocampal network, as well as WM intrinsic to cortical language regions in RH homologues, were associated with post-therapy outcomes. Variability in the functioning of these networks (indirectly inferred here through volume) may then give rise to variations in learning ability.

In a further study, Richter and colleagues (2008) delivered constraint induced aphasia therapy (CIAT) to 16 aphasic patients, and investigated activation patterns before and after treatment using task-based fMRI with reading and word stem completion tasks. Activation patterns prior to treatment predicted later success, and interestingly, these were in the RH, including the IFG/insula. The present results are in therefore in line with these findings which support a role of RH language networks in recovery. Furthermore, successful patients showed decreases in activation in these regions post treatment, whilst the least successful patients showed increases, suggesting the functioning of these networks varies amongst patients. Again, the variability in WM found in the present study, and its relation to treatment outcomes, are in line with these findings, which support the notion that variability in the functioning of RH networks relate to how well patients are able to engage with and learn from treatment.

In summary, the present study builds on previous work by providing convergent evidence, from structural imaging, which suggests that WM pathways in the RH are a predictor of treatment success. These findings are in agreement with previous studies which have generally shown a facilitatory role of activation in the RH in recovery, in the chronic phase of aphasia (Cocquyt et al., 2017).

5.3.3.6 Summary of cluster findings

When taken together, these clusters have been shown to align with regions involved in language specific processes, word learning networks, and further cognitive processes. Although these inferences are speculative, these are all processes which are likely to be recruited in a task such as Listen-In. Given the highly significant group results, what is known is that volume of WM confers a functional advantage in some way for patients during treatment. The correspondence of these clusters to overlapping neural regions implicated in both domain general and language specific functions suggests that better functioning WM networks in the RH support language treatment.

5.3.3.7 What accounts for variability in white matter volume?

A key question which remains is what accounts for variability of WM volume in the RH. One possibility is pre-morbid individual differences in brain structure and function. Although language has traditionally been regarded as left lateralised, evidence suggests that differences in lateralisation may be more prevalent amongst the population, and relate to recovery outcomes . For example, Catani and colleagues (2007) found that pathways between Broca's and Wernicke's areas were strongly left lateralised in 50% of healthy individuals, whilst being bilaterally distributed in 18% of individuals. In the present study, there is no way of knowing whether some patients had greater symmetry of language functions pre-stroke. Five out of twenty-five patients were left-handed pre-stroke which may increase the likelihood of differences; although only a minority of left-handed individuals have been shown to have RH language dominance (Knecht et al., 2000), subtle difference in hemispheric processing could still be present, as shown by

Catani's and colleagues (2007) results. In addition, speech comprehension has been argued to be bilaterally organised up to lexical recognition (Hickok & Poeppel, 2007), and it is possible that the degree of bilateral processing could also differ amongst patients and give rise to structural differences. For example, WM integrity in the right MTG in aphasic patients was found to predict later speech fluency outcomes, but when integrity was compared with controls, no differences were observed, suggestive of pre-morbid individual differences (Pani, Zheng, Wang, Norton, & Schlaug, 2016).

Given that all the patients were in the chronic phase, it is plausible that post-stroke reorganisation may also account for some of the variability observed in WM volume; however, this interpretation is speculative, as there is no way to distinguish the difference between less 'shrinkage' of neural matter (atrophy) and greater 'growth' of WM. In GM, better language performance has been associated with increased regional GM volume (compared to controls) in the right ATL, temporo-parietal cortex, and SMA in production (Hope et al., 2017; Lukic et al., 2017; Xing et al., 2016), and MTG in comprehension (Lukic et al., 2017), suggesting take over by homotopic regions. In line with this, some functional imaging studies have also demonstrated functional changes in the RH; however, some of these studies have shown detrimental effects of RH involvement on language functioning (Cocquyt, Ley, Santens, Borsel, & Letter, 2017). It is interesting to note the correspondence of RH clusters to regions where damage in the LH has been associated with speech comprehension impairments in aphasic patients. These include the anterior STG, MTG, STS and angular gyrus, and VLPFC (Bates et al., 2003; Dronkers et al., 2004). It has been suggested that patients with more severe aphasia may recruit undamaged RH homologues to support language processing (Crosson et al., 2007; Richter et al., 2008), so neural reorganisation may be particularly relevant to Listen-In patients, as they often had large and extensive lesions to these key LH language regions. For example, in a study with central alexia stroke patients, average lesion volume was 163cm^3 (Woodhead et al., 2018), whilst the average lesion volume in the present sample was 216cm^3, suggesting the present sample may represent particularly severe patients. It is unclear whether WM changes, to the extent and breadth of clusters found in the present study, could be due to these mechanisms, as previously, WM changes have been correlated with specific skills and localised to small discrete brain regions. For example, microstructural changes have

been observed in the RH in response to intensive Melodic Intonation Therapy, in regions around the arcuate fasciculous (Wan et al., 2014).

Another possibility, not mutually exclusive, is that declines in WM volume account for variability. In normal ageing there are declines in both GM and WM (Draganski et al., 2011; Good et al., 2001; Sexton et al., 2014). Stroke patients are more likely to show cognitive declines, and are also more likely to develop dementia, and show signs of WM damage (WM hyperintensities) which have been associated with cognitive impairment (Burton et al., 2004; Stebbins et al., 2008). Outside of the lesion region, atrophy in stroke patients also occurs faster than in normal ageing (Seghier et al., 2014), and increases as well as decreases in structure have been linked to changes in language skills in chronic aphasia (Hope et al., 2017). A combination of normal ageing and neurobiological sequalae of stroke (such as Wallerian degeneration and transcallosal degeneration) may therefore give rise to differences in WM volume, and a functional outcome of this may be differences in cognitive performance and learning ability during treatment. As evident from the scatterplot, many patients show homogenous WM volumes suggestive of a normal range, whilst several patients show relatively reduced volumes in comparison, which could relate to accelerated atrophy. Of note is that age and time since stroke were accounted for in the present model, so if declines in WM do contribute to variability, they are unlikely to be due to systematic age-related changes across participants. It is possible that atrophy is present in a non-linear fashion amongst patients, or that a combination of the factors described may give rise to differences in WM volume across participants.

The cross-sectional design of this study means it is not possible to adjudicate between these possibilities. It is therefore not clear whether variability in volume is different than would be expected in a healthy age matched cohort. As highlighted, this is an important consideration as it would enable inferences to be made about post-stroke structural adaption in the RH (hypertrophy), or equally, atrophy, and how this relates to treatment outcomes. Future work may look at comparing volume of WM in regions of interest in the current patients, with a group of age-matched controls (healthy, and non-aphasic stroke) to help determine between these possibilities (as in Xing et al., 2016). However, of note is that Pani and colleagues (2016) found correlations between structure and performance

in the RH their aphasic group, but found no overall differences in structure when compared with a control group, therefore null findings may need to be interpreted with caution.

5.3.3.8 Limitations

A limitation of this study is the small sample size (N=25) which increases the likelihood of false positive and negative results in this type of mass-univariate approach (Lorca-Puls et al., 2018). Difficulties replicating structural brain-behaviour correlations in healthy individuals have been reported (Boekel et al., 2015). In patient studies, replication may be further impacted by increased heterogeneity between samples in characteristics such as age, lesion size and site, and fatigue. For example, the current study did not exclude patients with a history of previous stroke as it sought to reflect a typical clinical caseload of patients; however, this is a common exclusion criterion in many aphasia studies. As a result, three patients had previous lesions in their RH. In this way, the current finding could be attributed to sample specific effects, therefore caution is needed in generalising findings to the aphasia population. For example, Gajardo-Vidal and colleagues (2018) have recently demonstrated that lesion-symptom mapping results can be driven by subsets of patients, with subsequent iterations changing the location of significant regions. Examination of scatterplots in the present study demonstrates a clear trend in the sample which doesn't appear to be driven by only these patients; however, it is possible that they provided sufficient variability to reveal a significant effect which would otherwise have remained below the significance threshold in the VBM analyses. Nevertheless, given the minimal inclusion and exclusion criteria of the current study, it could be argued that these findings are more easily generalisable to patients with impaired speech comprehension, than studies where patients have been selected based on a specific set of behavioural criteria.

A further limitation common to all VBM approaches is interpreting the measure of tissue volume at a cellular level. This approach is based on tissue probability maps, which give a probability that each voxel belongs to a particular tissue class. In the present study,

modulation was also used which encodes volume of the voxel prior to normalisation. Therefore, "volume" in the present study refers to both tissue probability and volume. It is therefore not clear how volume of GM and WM in the present findings relate to underlying neural architecture. In addition, Lövdén and colleagues (2013) point out considerable variability in reporting of labels and units of measurements with this approach, and call for greater consistency to be used. The present study aimed to accord with this by following standard procedures and terminology as used by the developers of this approach (Ashburner & Friston, 2000). In sum, in is not clear how volume relates to the microstructure of GM and WM. Nevertheless, this method has provided information which shows that variability in tissue composition, regardless of the underlying cause, relates to behavioural outcomes, and therefore must relate to functioning of the structures which it supports.

5.3.3.9 Future directions

Patients who entered treatment with greater volume in predominantly RH WM networks were at an advantage for responding to treatment. The location of clusters showed correspondence to structures which have been functionally implicated in a range of cognitive and language tasks. Although these findings were structural, they correlated with a measure which represented learning success, suggesting greater volume (or reduced degeneration) in the right hemisphere is associated with an increased ability to learn new words.

Previous evidence has suggested that aphasic patients adaptively recruit RH regions during recovery. The present findings lend some support to this view, as volume of WM in mainly the RH was a marker of treatment success. Alternatively, it may be that patients recruit the RH in a typical manner, and that reduced degeneration (relative to other patients in the sample) confers some advantage during treatment. This may equally account for the correlation between WM volume and response to treatment in the present findings. It is not possible to differentiate between these explanations. Further investigation is needed to account for the variability observed in WM volume, and this

may be achieved by comparing cluster volumes with healthy controls, and non-aphasic stroke controls.

Given the novelty of these findings, and recommendation by Gajardo-Vidal and colleagues (2018) that this approach cannot be used to predict outcomes in other patients, further evidence is needed before firm conclusions can be drawn about the predictive role of RH WM structure in treatment related recovery. For these reasons, the clinical implications for these findings are tentative. In the advent of patients increasingly undergoing routine MRI scans post-stroke, imaging may be one way that clinicians can better understand individual patient profiles and direct personalised treatments. For example, better understanding of RH structure and function, and relation to treatments, may aid in developing treatments which build on residual strengths.

One hypothesis from this study is that WM clusters represent networks supporting a range of linguistic and non-linguistic cognitive processes which support treatment. A future direction for Listen-In could be to adapt the type of treatment task to decrease cognitive demands. A different approach, such as errorless learning, may be more beneficial for some patients who have reduced cognitive skills. For example, Fillingham and colleagues (2006) found that patients with better working memory, recall memory and attention performed better in an errorful treatment (like Listen-In), and may even demonstrate enhanced learning. Likewise, an errorless learning technique may be more beneficial for patients less able to draw on executive functions. In this case, patients would be provided with the auditory stimulus and correct picture, and would always make a correct choice. Errorful and errorless treatment paradigms have tended to show similar levels of efficacy in anomia treatment (Fillingham et al., 2003), but no such comparison has been made for speech comprehension treatments, therefore an investigation into these different approaches would first be warranted.

5.4 Chapter 4: Investigating structural brain adaptation in response to Listen-In treatment

Aim

To investigate whether Listen-In induces localised changes in brain tissue, in patients with chronic aphasia.

Hypotheses

(1) Listen-In therapy will induce relative increases in grey and/or white matter volumes in key speech processing networks in the left hemisphere.
(2) Listen-In therapy will induce relative increases in grey and/or white matter volumes in key speech processing networks in the right hemisphere.
(3) Total exposure to auditory stimuli over the Listen-In treatment block will correlate with tissue changes in bilateral temporal lobes, in the speech processing network.

5.4.1 Experimental Procedures

5.4.1.1 Participants

Twenty-five patients data were included in the present anlaysis, for the same subgroup (N=25) of patients reported in Chapter 3.

5.4.1.2 Data preprocessing

Images were smoothed with an isotropic kernel of 6mm full-width half-maximum. For prior preprocessing steps see Methods.

5.4.1.1 Data analyses

Twenty-five smoothed GM and WM change images were then entered into four separate simple linear regression models in SPM12 (Figure 5-23). Change images represent change in concentration of GM and WM over the treatment period, more than a within-subject control period (standard care) (see Methods section for processing steps). Analyses 1 and 2 were designed to identify regions where change in concentration of GM and WM correlated with Listen-In treatment response. Percentage improvement on trained items on the ACT, from pre to post therapy, was used as the dependent variable, as this was the significant behavioural effect from Chapter 2. Analyses 3 and 4 were designed to identify regions where change in concentration of GM and WM correlated with total time spent on treatment. This was entered as total number of hours spent on Listen-In therapy. Between subjects effects were accounted for in the within-subject design during pre-processing steps, therefore no further regressors were entered into the model. In keeping with standard practice, the statistical voxel-level threshold was set at $p< 0.001$ uncorrected with cluster-level significance set at $p < 0.05$ after family-wise error (FWE) correction for multiple comparisons. For analyses 1 and 2, the analysis was restricted to bilateral STG, which was anatomically defined using the WFU PickAtlas toolbox in SPM12 (Maldjian, Laurienti, Kraft, & Burdette, 2003). This mask was used to perform a small volume correction in SPM. This anatomical prior was based on the following two observations: in healthy individuals, meta-analyses of functional imaging studies show activation of bilateral STG during speech processing tasks (DeWitt & Rauschecker, 2012; Vigneau et al., 2011); and in aphasic patients, functional changes in bilateral STG have been found in response to auditory training (Musso et al., 1999; Z. V. Woodhead et al., 2017).

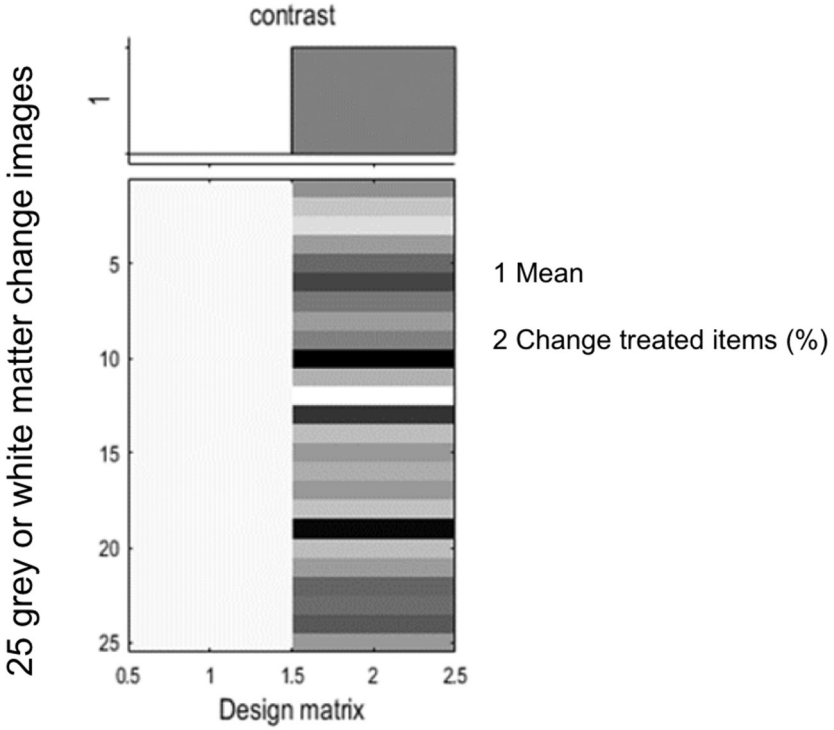


Figure 5-23 Design matrix used in simple linear regression models in SPM12. All four models had one regressor of interest: change in treated item performance (%), and time spent on therapy (hours).

5.4.2 Results

5.4.2.1 Analyses 1 and 2: Relation between Listen-In treatment response, and change in tissue density

Analyses 1 and 2 idenified regions were change in GM and WM density covaried with change on treated item performance from pre to post treatment. For GM, small volume correction resulted in one significant cluster located in the posterior part of the right STG, on the posterior inferior boundary of the planum temporal (Figure 5-24, Table 5-8). For WM, small volume correction revealed one significant cluster located in the mid part of the left STG, bordering the STS (Figure 5-24, Table 5-8). Figure 5-27 displays these clusters on an average GM and WM template from 25 patients. The scatterplots in Figure 5-24 illustrate the relation between change in GM and WM and treatment response for individual patients.

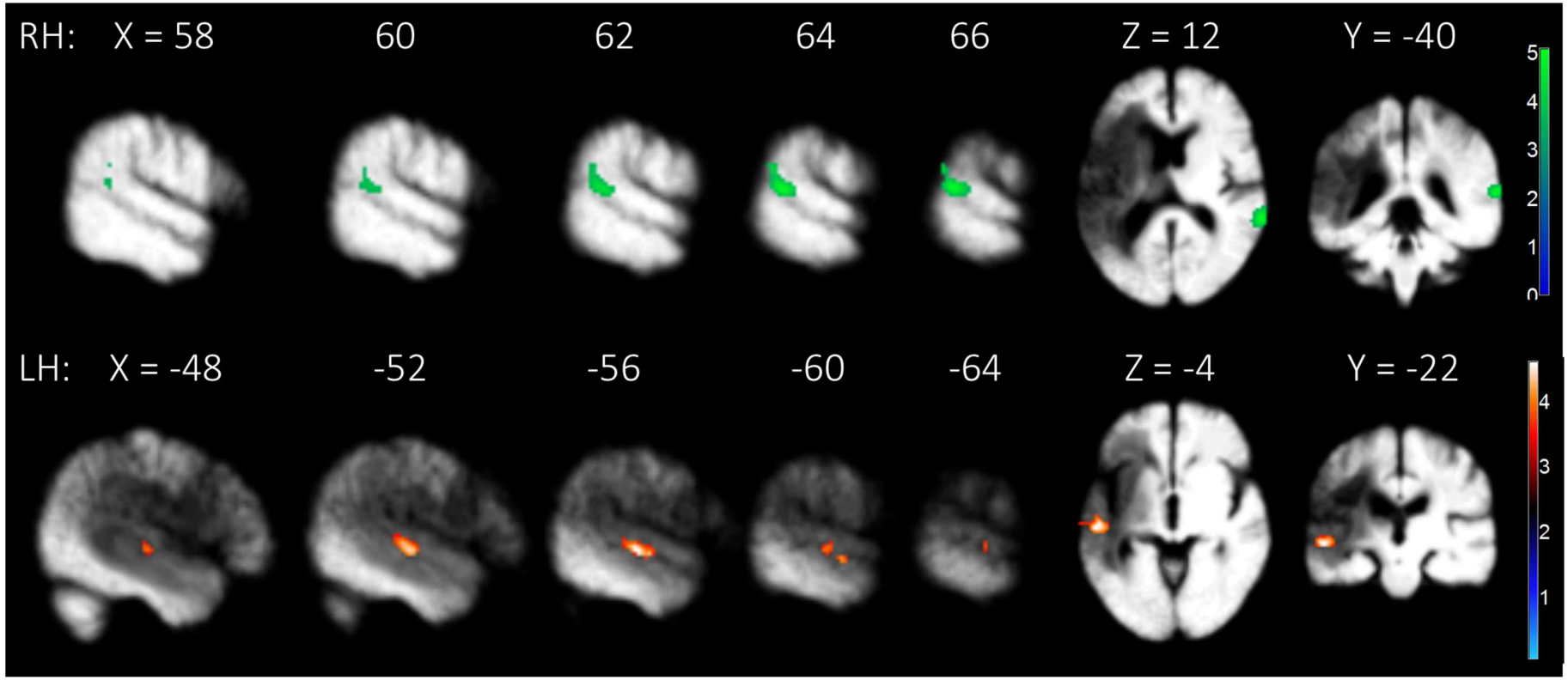


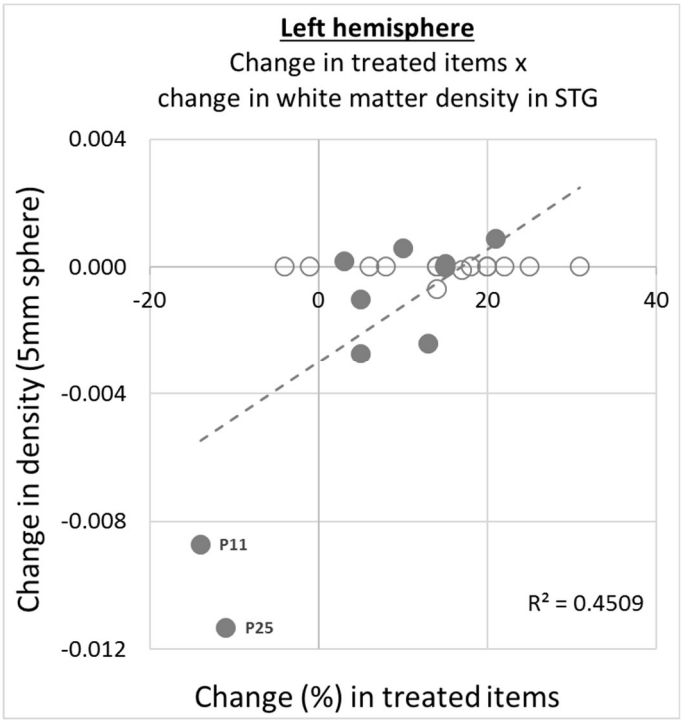
Figure 5-24 Clusters where change in tissue density significantly covaried with change in treated item performance from pre to post treatment, for grey matter (blue) and white matter (orange). Colour bars represent t-values at that voxel

	Cluster size (k_E)	T - value	Peak MNI coordinates (mm)			Brain structure/function
			x	y	z	
Analyses 1 and 2: Simple linear regression with treatment response						
Left WM	109	4.59	-56	-22	-4	STG (BA22)
Previous findings:						
DeWitt & Rauschecker (2012)			-58	-18	0[1]	Phonological speech processing (healthy)
Mårtensson et al. (2012)			-59	-11	2	Structural plasticity related to foreign language proficiency (healthy)
Right GM	154	5.05	66	-40	12	pSTG (BA22)
Previous findings:						
Xing et al. (2016)			66	-43	15	GM volume related to spontaneous speech and repetition ability (aphasic)
			60	-45	23	
Analyses 3 and 4: Simple linear regression with treatment dose						
Left WM	581	7.39	-36	-52	12	Posterior thalamic radiation
		6.79	-42	-34	-10	Inferior longitudinal fasciculous/occipito-frontal fasciculous

Table 5-8 Significant clusters from longitudinal voxel-based morphometry analysis for treatment response (Analyses 1 and 2) and dose (Analyses 3 and 4). Voxel-wise threshold set at p<.001, FWE correction at p<.05. For Analyses 1 and 2, FWE adjusted for small volume correction in bilateral STG. Coordinates are the peak voxel with the highest t-score in that cluster. GM=grey matter, WM=white matter

[1] Coordinates converted from Talairach to MNI for comparison purposes, using the same ICBM2TAL transformation used by DeWitt and Rauschecker (2012), accessed from: http://brainmap.org/icbm2tal/. Original Talairach coordinates: [-58, -20, 2].

(A)



(B)

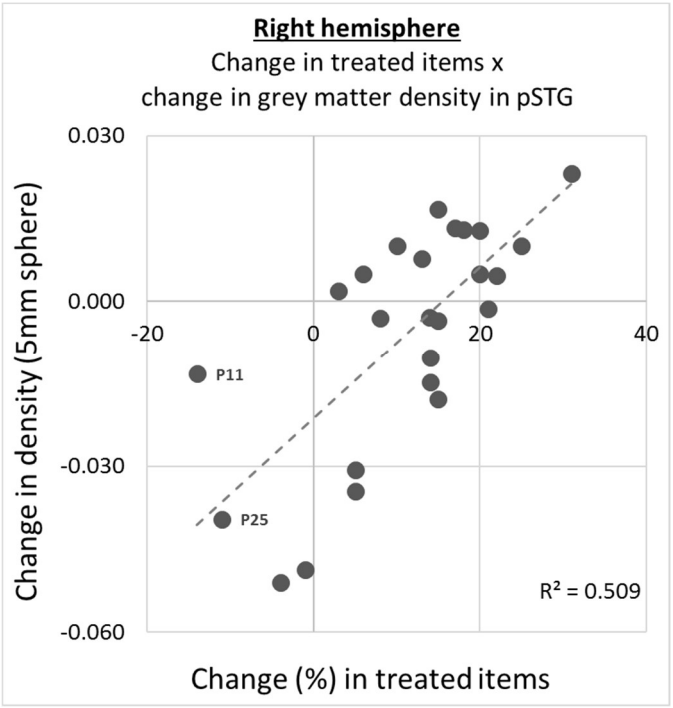


Figure 5-25 Scatterplots showing the relation between change in (A) WM and (B) GM density, and change in treated item performance over the treatment block, for 25 patients. In the left hemisphere, unfilled circles represent patients with a lesion to the peak voxel in that cluster. Density values are eigenvalues from 5mm spheres around the peak voxel in that cluster.

5.4.2.2 *Analyses 3 and 4: Relation between Listen-In dose (hours), and change in tissue density*

Analyses 3 and 4 identified areas where change in GM and WM density covaried with total time spent on Listen-In therapy challenges over the 12-week treatment (as this varied between subjects (M=85, SD=31). In Analysis 3, whole brain analysis resulted in the identification of one significant GM cluster located in deep white matter corresponding to the sagittal stratum (Table 5-8). Figure 5-26 displays this cluster overlaid on the JHU white matter tractography atlas in MRIcron. The two highest peaks correspond to the posterior thalamic radiation, and the inferior longitudinal fasciculous/occipito-frontal fasciculous, respectively. The scatteplot in Figure 5-27 shows the relation between change in tissue density and treatment dose for individual patients. Note that the second highest peak was chosen to illustrate this association due to greater variability, as fewer patients show lesions to this region of the cluster. In Analysis 4, whole brain WM analysis revelaed no regions where density change correlated with treatment dose.

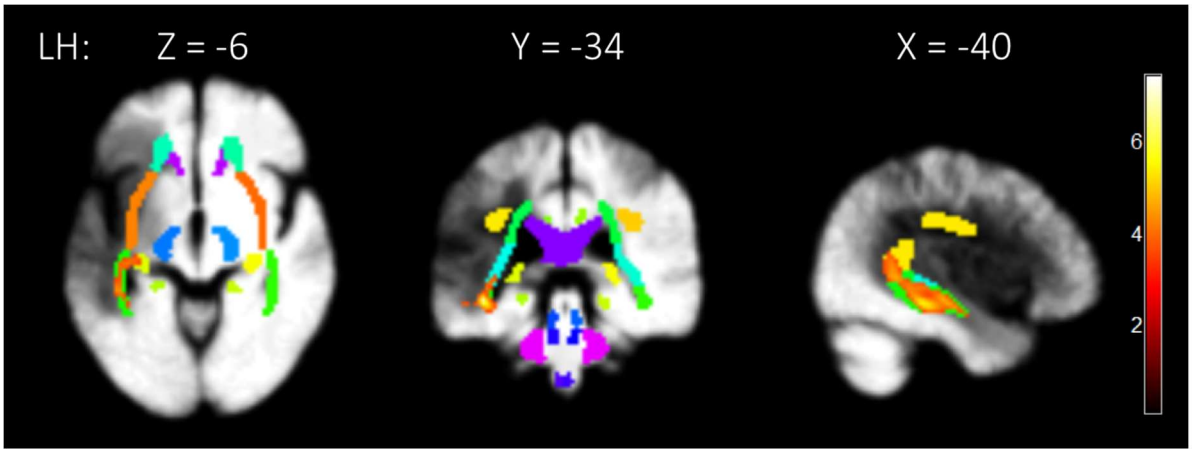


Figure 5-26 Cluster where change in density covaried with hours spent on therapy challenges (in heatmap colours, see bar), overlaid on the white matter tractography atlas within MRIcron. Brain template is an average GM and WM image for 25 patients. The colour bar indicates the t-value at that voxel.

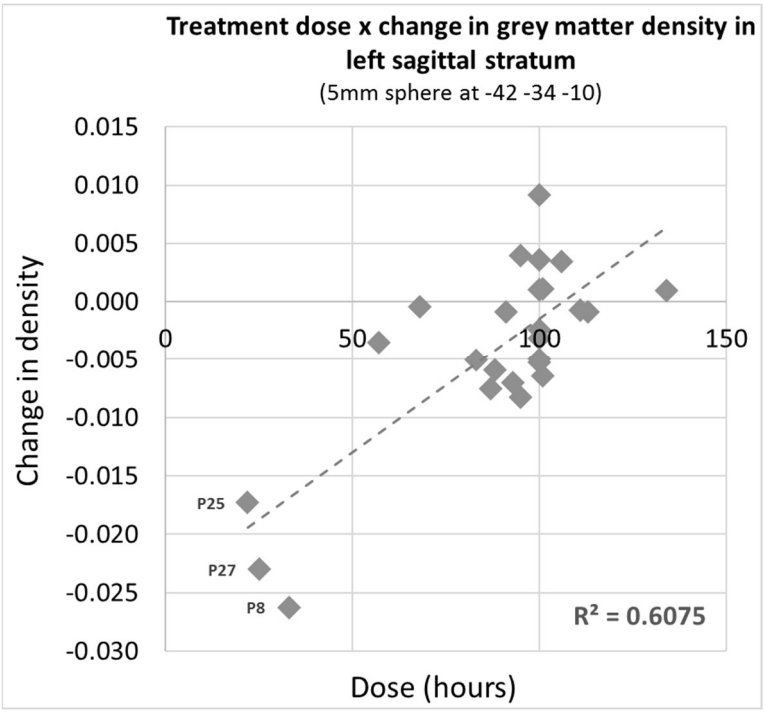


Figure 5-27 Scatterplot showing the relation between change in tissue density and treatment dose in hours, for 25 patients. Density change is an eigenvariate from a 5mm sphere around the second highest peak in that cluster.

5.4.2.3 Analysis 5: Relation between baseline white matter volume, and change in GM density

In Chapter 2, volume of tissue in five regions in the RH, prior to treatment, was related to improvements in spoken word comprehension post-treatment. In the present chapter, these behavioural improvements were related to changes in GM tissue density in the right pSTG. One question is whether there is a relation between patients baseline WM volume in the RH, and change in GM tissue density post-treatment. In other words, did patients with greater volume of WM, prior to treatment, go on to show the greatest changes in the pSTG.

To investigate this question, a group wise analysis was conducted which correlated patients WM volume with change in GM density. WM volume for each patient was total WM volume (ml^3) at baseline, in the significant clusters identified in Chapter 2. These were calculated using a binary mask of significant clusters, applied to patients segmented and modulated WM images, as described previously. The results of this correlation are

presented in the scatterplot in Figure 5-28. Spearman's rank-order correlation showed a moderate positive association between these two measures, which was significant (r_s=.412, p=.04). This indicates that patients with greater WM volume at baseline within the RH clusters, tended to show greater changes in GM tissue in the pSTG in response to treatment.

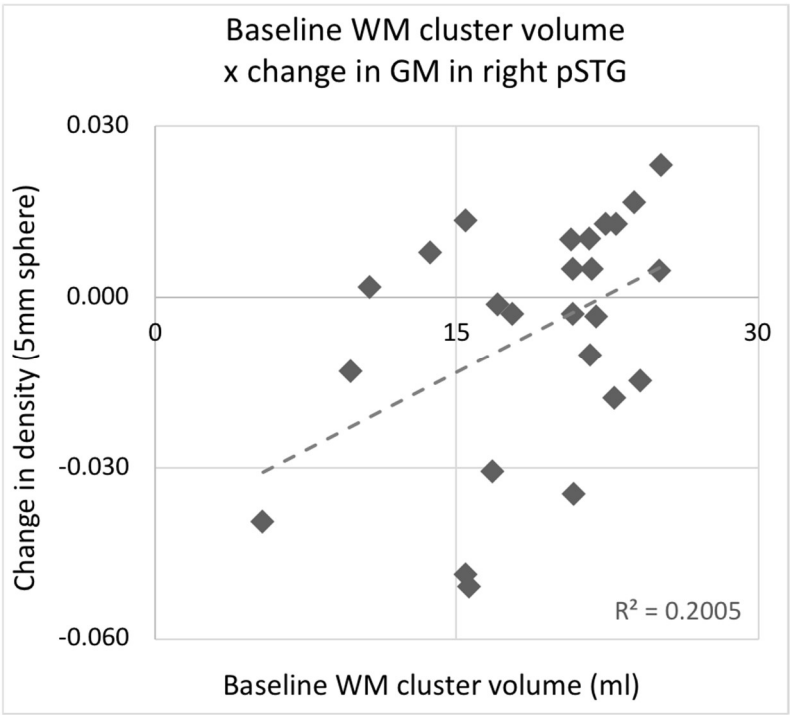


Figure 5-28 Scatterplot showing the relation between regional WM volume in the right hemisphere at baseline, and treatment induced change in GM in the right pSTG. Change in density is an eigenvariate from a 5mm sphere around the peak voxel in the right pSTG cluster

5.4.3 Discussion

In Hypothesis 1 and 2, it was predicted that Listen-In therapy would induce relative increases in grey and/or white matter volumes in key speech processing networks in the left and right hemispheres. Both of these hypotheses were supported. Changes in tissue were found to covary with change in treated item performance in two key regions of the speech comprehension network: in the LH, this was in WM in the mid-STG, a key part of the ventral speech processing stream (Analysis 2); and in the RH, this was in the posterior STG, a homologue to Wernicke's area in the LH (DeWitt & Rauschecker, 2012; Scott et al., 2000). Crucially, both of these structural changes correlated with changes in treated item performance, suggesting specificity of these changes to Listen-In treatment.

5.4.3.1 Comparison with experience-dependent neuroplasticity findings in studies of human expert performance, and therapeutic studies in stroke patients

In healthy individuals, a number of studies have demonstrated experience-dependent changes in GM and WM in response to different types of training tasks (Bezzola et al., 2011; Draganski et al., 2004; Hamzei et al., 2012; Ilg et al., 2008; Lövdén et al., 2013; Scholz et al., 2009). A small number of these studies have demonstrated changes within the language domain (Mårtensson et al., 2012; Schlegel et al., 2012; Schmidt-Wilcke et al., 2010). However, many of these studies have used cross-sectional designs, and so longitudinal changes have been reported less frequently. Furthermore, in their review of the experience-dependent neuroplasticity literature, Thomas & Baker (2013) report only seven studies which correlated structural changes with behavioural measures. The present study builds on these previous findings by offering new longitudinal evidence for neuroplasticity in the speech processing domain, which directly relates to treatment outcomes in aphasic patients.

Previously, studies have suggested neural adaptation in aphasic patients in relation to gradual and spontaneous recovery in chronic aphasia (Hope et al., 2017; Xing et al., 2016). The current findings add to this evidence by demonstrating neural changes in relation to

short term treatment over 12-weeks. Few previous studies have reported longitudinal structural changes in aphasic patients in response to treatment. In two studies which did report structural findings, changes in WM integrity were observed. In contrast to VBM, these studies utilised diffusion tensor imaging (DTI), which is specialised for detecting subtle variations in WM. In the first study, Schlaug, Marchina and Norton (2009) investigated WM networks before and after 75 sessions of Melodic Intonation Therapy (MIT) in six patients with chronic aphasia. DTI showed an increase in fibres and volume in the arcuate fasciculous (a key fibre tract linking fronto-temporal language regions) following treatment; however, there was no direct relation to behavioural treatment outcomes, so the finding is hard to link with performance-related improvements in behaviour. There was also no control group, and a small number of patients. In a second study with eleven patients and a non-treatment control group, Wan and colleagues (2014) also gave MIT to their patients, and found reductions in FA in the right IFG, pSTG and posterior cingulum, from pre to post treatment. Speech production improvements were also correlated with reductions in FA in the right IFG. This second study provides much stronger support for experience induced changes in WM, and shows that modifications to WM architecture can be induced by practice-based treatment in aphasic patients. As in the present study, changes were observed in biologically plausible regions which related to the treatment task.

The present findings are in also line with studies demonstrating changes in activation patterns in response to treatment. Both increases and decreases in activation patterns have been found. For example, Marcotte and colleagues (2018) administered a phonological treatment to two chronic aphasic patients, and found task-based activation (naming of treated words) increased and decreased in different brain regions, for both patients. These findings are reflective of the heterogeneity amongst patients in functional activation patterns during and following treatment (Kiran & Thompson, 2019). In speech comprehension, few studies have investigated functional changes in response to treatment. In one study which did, phonological training resulted in inter-hemispheric changes in activity in the pSTG which were associated with treatment success (Z. V. Woodhead et al., 2017). A second study found an increase in pSTG activity in the RH which correlated with sentence comprehension improvements following treatment (Musso et

al., 1999). Although a clear picture has not yet emerged, these functional imaging findings demonstrate the ability of the brain to adapt, and repeated activation of these mechanisms may induce longer term structural changes. The present findings provide convergent evidence which support this prediction, as highly localised structural changes were observed following a specific and repetitive treatment. Furthermore, these changes were in key speech comprehension regions, in line with the findings of Woodhead and colleagues (2017) and Musso and colleagues (1999).

To my knowledge, this is the first study to show therapy induced changes in brain tissue in response to speech comprehension treatment, therefore these findings offer a novel contribution to the field of aphasia rehabilitation by showing that aphasic patients, like healthy individuals, can demonstrate experience-dependent structural changes, and that targeted treatment can induce these changes. These findings also corroborate those of Buchel and colleagues (2008) by showing that experience dependent neuroplasticity mechanisms are maintained, at least to some extent, in older age. Given the novelty of these findings, some caution is needed, particularly as few studies have reported on this particular method (Ashburner, 2013). Nevertheless, the presence of changes in biologically plausible regions, which relate to the treatment task, supports the validity of these findings.

5.4.3.2 Function of the left hemisphere cluster

The cluster in the LH corresponds to WM underlying the mid-STG, in Brodmann's area 22. The peak coordinate in this cluster aligns closely to the mid-STG region, which has been implicated in processing phonetic aspects of speech across a wide number of studies in healthy individuals (DeWitt & Rauschecker, 2012) (see Table 5-8). These functional imaging studies involved contrasting speech stimuli (such as consonant-vowel syllables and pseudowords) with matched sounds (such as tones, noise and degraded speech) to isolate regions specific to the acoustic phonetic components of speech. The correspondence between the present cluster and its function in healthy speech processing is in line with the demands of the Listen-In task. As well as listening to speech, all therapy trials included phonological distractors, so patients needed to attend to the

phonetic components of target words in order to discriminate between them (e.g. tie/pie). Taken in this context, one possibility is that changes in WM tissue in this region relate to recruitment of this region during phonetic aspects of speech processing.

In support of this account, tissue changes in this region have also been identified in individuals learning a foreign language. Mårtensson and colleagues (2012) investigated a group of interpreters on an intensive three-month language training program, where students were tasked with learning 300-500 new words each week. Individuals who acquired greater proficiency in the foreign language at the end of the course demonstrated greater increases in cortical thickness in the left mid-posterior STG, as well as changes in volume in the right hippocampus. The coordinates of this cluster align closely with the peak coordinates identified in the current aphasic patients who were also tasked with learning, or-relearning, phonological-semantic associations (Table 5-8). The two tasks were therefore similar in nature, and provide complimentary findings.

However, some caution is warranted with this interpretation. The scatterplot in Figure 5-25 shows that more than half of patients had lesion damage to the mid-STG, and accordingly, these patients show no changes in WM density. The correlation appears to be driven mainly by the ten remaining participants, and for this reason, inferences regarding neuroplasticity here can only be inferred for this small subgroup of patients, and are not representative of the full cohort. Of the remaining ten patients, two appear to be significant outliers, showing much greater decreases in WM density, and are the two worst performing patients overall, showing declines in comprehension of treated items post therapy. The large difference in density change compared to the more homogenous range seen in the further eight patients suggests some form of accelerated atrophy in these two patients, which was detrimental for responding to Listen-In treatment. Previous investigations into VBM methodology have demonstrated that statistically significant results can be driven by small subsets of patients (Gajardo-Vidal et al., 2018), therefore caution should be taken with this result. Nevertheless, the further eight patients show variability in tissue change, and generally conform to a positive association between tissue change and behavioural outcomes, although this correlation is weaker. The correspondence of these findings to Mårtensson and colleagues (2012) interpreters also

adds support to the interpretation that these may be true treatment effects. With these considerations in mind, assumptions regarding neuroplasticity in this region due to treatment are tentative. Given two patients are driving these findings, it could be that decline in density (whether local or more global in nature) precluded these two patients from responding to treatment, accounting for this correlation, rather than a genuine linear relation between treatment success and plasticity. This is still an interesting interpretation as it would suggest that this region is key for this type of task.

5.4.3.3 Function of the right hemisphere cluster

In the right hemisphere, changes in GM tissue were observed in the pSTG, almost exclusively within Brodmann's area 22, a RH homologue of Wernicke's area. In dual route models, posterior parts of the STG correspond to the dorsal language pathway and arcuate fasciculous (Catani & Mesulam, 2008; Hickok & Poeppel, 2007; Saur et al., 2008). The pSTG in particular has been put forward as a central node in phonological processing, important for the early stages of mapping sound to meaning (Saur et al., 2010). Saur and colleagues (2010) report the presence of identical but weaker networks in the RH for phonological processing; however, the nature and extent of the RHs contribution is still a matter of debate. Some have put forward models which propose bilateral and parallel streams of speech processing in both hemispheres (Hickok & Poeppel, 2007), and this has been supported by a wealth of studies in healthy individuals which show speech processing activates bilateral networks (Vigneau et al., 2011). In their meta-analysis, Vigneau and colleagues (2011) found two regions in the right temporal lobe associated with phonological processing: Heshl's Gyrus, and the pSTS. Activations in the pSTS (close the present cluster) occurred in tasks which contrasted speech sounds (e.g. pseudowords, syllables, vocal sounds) with non-speech sounds (e.g. tones, noise).

These findings demonstrate that the RH is functionally recruited during speech processing tasks in healthy individuals, and therefore in Listen-In, repeated exposure to auditory speech stimuli would be expected to recruit these bilateral networks, albeit with likely individual differences related to lesion damage. In line with this, change in GM density in the right pSTG was observed in response to treatment. In light of the previous findings,

one interpretation is that these changes reflect recruitment of typical RH regions, supporting the hypothesis that patients recruit typical networks in the RH to support speech processing. Given the clusters posterior position, this is likely to correspond to a dorsal pathway, which projects dorsoposteriorly away from the STG (Hickok & Poeppel, 2007; Saur et al., 2008). The precise function of RH regions in speech processing is not yet clear, however, one hypothesis is that dorsal processing in the RH may represent speech as a non-categorical acoustic signal (versus speech specific phoneme representations) (Hickok & Poeppel, 2007), allowing the RH to compensate for LH damage by providing an alternative, but perhaps less optimal, route for speech processing. In this way, patients in Listen-In who were better able to compensate via RH mechanisms may have been at an advantage during treatment, accounting for the correlation between improvement in treated item performance and tissue change in this region. This could relate to premorbid differences in RH speech processing networks, or compensatory post-stroke mechanisms.

However, one caveat to this interpretation is that the region identified here lies in a more posterior position to regions previously implicated speech listening tasks, which in the posterior temporal lobe, are typically within the STS (Vigneau et al., 2011). It is important to note that Listen-In wasn't a controlled listening task, and so the demands of the task are different from those typically reported in healthy speech processing studies, which may account for these discrepancies. For example, as well as listening to the stimuli, patients were required to maintain the target in verbal short-term memory, and decide between a range of distractor items. It is highly likely that other strategies are likely to have come into play, such as rehearsal and repetition, to facilitate the task, which may engage alternative regions.

Interestingly, the peak voxel in this cluster in the right pSTG region very closely overlaps with an area of RH adaptation which has been previously identified in aphasic patients. In their cross-sectional study, Xing and colleagues (2016) used VBM to correlate language skills in chronic aphasic patients, with GM volume in the RH. They identified several clusters where GM volume positively correlated language skills, and these related to repetition, naming/word retrieval, and spontaneous speech; furthermore, many of these clusters correlated with digit span and pseudoword repetition ability, suggesting that in

their patients, these regions in the RH related to wider verbal working memory and motor speech output processes. Of note is that the present cluster corresponds directly with the spontaneous speech cluster, and partially with the repetition cluster identified in these aphasic patients (see Table 5-8). Conversely, however, in the present study, this particular region in the pSTG was related to improvements in speech comprehension. Although Listen-In was training speech comprehension, the behavioural variable was improvement on treated items. Therefore, this region of structural change could relate to strategies commonly recruited by patients to support performance on the task, rather than speech comprehension per se. In line with the postulated function of this region in Xing and colleagues (2016) patients, repetition and rehearsal may be likely candidates in Listen-In patients. These two strategies are likely to have been particularly important during spoken word-to-picture matching, to maintain the word or sentence in verbal short-term memory, whilst choosing a response.

Xing and colleagues' (2016) patients also showed greater GM volume in RH regions than a group of both healthy, and stroke, control patients, leading the authors to speculate that these regions may have undergone post-stroke hypertrophy, driven by language experience during recovery. This finding is in line with a study by Leff and colleagues (2002), who set out to investigate whether a laterality shift could be observed in patients with damage to the left pSTS, who had recovered single word comprehension. Following an asymmetrical response in the LH in response to listening to words in healthy participants (measured using PET), the same task was given to aphasic patients. Like controls, patients showed the same mean activity in the pSTS, but in the RH, and they also showed steeper activity in the RH than both healthy control and patients with a non-lesioned left pSTS. The authors suggest that increased activity, distinct to this patient group, represents reorganisation in the homotopic region. As these patients had recovered single word processing, one hypothesis is that homotopic compensation took place. The present finding adds significant support to this hypothesis by revealing direct evidence of structural neuroplasticity in a similar area in the right pSTG in response to speech comprehension treatment. Furthermore, changes here were related to a specific treatment, rather than generalised recovery as in the previous two studies, suggesting areas of recruitment common to both spontaneous and treatment induced recovery.

A pertinent question is whether the region identified here in the pSTG relates to typical RH functions, or to compensatory reorganisation in an alternative region. As previously noted, this region in the STG is marginally more posterior than typical RH activations during speech processing (Vigneau et al., 2011). Previously, Meltzer and colleagues (2013) investigated sentence comprehension in aphasic patients using magnetoencephalography (MEG). They found activation in bilateral posterior temporal and parietal regions predicted better sentence comprehension performance, and that during a delay period, performance correlated with activity in fronto-parietal regions, regions which weren't active in controls. The authors speculate that this activity reflects alternative compensatory strategies, such as short-term memory, whereby patients rely on "effortful reprocessing". It is interesting to note that the present study also identified a posterior temporal region, therefore one possibility, in line with Meltzer and colleague's (2013) interpretation, is that Listen-In patients recruited an alternative region related to reprocessing strategies. This relates to the previous hypothesis that patients engage strategies such as rehearsal to support their comprehension.

It is not possible to adjudicate between these possibilities based on the present data. However, it is interesting to note that a measure of task success, but not a measure of task exposure (dose), was associated with tissue change in this region. If neuroplasticity was related to initial processing of auditory input, as it could be argued to be in this posterior STG region, then time spent listening to therapy challenges might have been expected to correlate with neural changes. Instead, patients who tended to do better behaviourally, tended to show greater increases in density (or lesser atrophy). This suggests a function beyond perceptual speech processes.

The GM scatterplot shows a relatively even distribution of changes in GM across patients, however, these relate to both increases and decreases in GM tissue density. As previously described, experience-dependent neuroplasticity findings in healthy individuals are typically characterised by increases in GM in response to training tasks. However, participants in these studies are typically young, healthy individuals. Conversely, Listen-In patients have large lesions, and are likely to present with neurobiological changes

associated with ageing and the sequalae of stroke. Stroke patients show accelerated rates of atrophy (~.95% per year), faster than typical age-related atrophy (~.5%) (Seghier et al., 2014), and so structural changes should be considered in relation to this specific population. Seghier and colleagues (2014) found shrinkage was unrelated to language recovery, and hypothesised it may relate to both degenerative and restorative processes. One possibility is that Listen-In training induced local increases in tissue density, but in the context of generalised atrophy, resulted in a net decrease in tissue density. In other words, global atrophy (degeneration) may have masked local increases in volume relating to treatment (restorative), so that lesser atrophy thereby reflects a treatment effect. A second possibility is simply that Listen-In slowed atrophy by stimulating the language network. Either interpretation may account for why a linear association was found which spanned both negative and positive values of density change, showing that for many patients, increases in density were related to training success. It therefore seems unlikely that patients with negative density values present with qualitatively different neuroplasticity mechanisms.

As demonstrated by the present findings, VBM is a sensitive measure which can detect morphological changes relating to the macrostructure of the brain. At a microstructural level, however, there is lack of specificity about what voxel values represent. In GM, the most likely candidates in this cortical region (STG) are changes in the number of non-neuronal cells (gliogenesis), changes in the number of neuronal synapses (synaptogenesis), changes in dendritic spine morphology, and changes in vasculature (angiogenesis) (Zatorre et al., 2012). In WM, changes may relate to axons (branching, sprouting, number, density and diameter), their myelination (new myelination or changes to myelin thickness), changes in astrocytes (morphology and number), and angiogenesis (Zatorre et al., 2012). Changes are therefore likely to reflect remodelling of neuronal processes, rather than the proliferation of new neurons, as in neurogenesis (Zatorre et al., 2012). These remodelling processes can be observed with imaging techniques due to changes in the ratio of cellular and extracellular space, which lead to differences in signal intensity at a particular voxel. In the present study, changes in concentration may therefore represent any number of these processes. As it is not possible to identify what

kind of histological processes underlie these changes, inferences regarding neuroplasticity are speculative.

5.4.3.4 Association between treatment dose and change in tissue density

In Analyses 3, time spent on therapy challenges positively covaried with change in tissue volumes in the LH, in a cluster which corresponded to WM pathways in the temporal lobe, showing overlap mainly with the ILF, and IFOF. These findings provide further support for Hypothesis 1, as they show changes in WM pathways underlying the ventral speech processing pathway (Almairac et al., 2015), although these changes correlated with exposure to the task, rather than treatment response. It is of note that this cluster was identified in the GM analysis; however, a comparison with WM tracts, using a well-established WM atlas (Oishi et al., 2008), revealed precise overlap with WM pathways in the temporal lobe, and therefore likely reflects change in WM rather than GM. During pre-processing, T1 structural scans were segmented into GM and WM which produced tissue probability maps, which give a proportion likelihood for a particular tissue class at each voxel. This cluster may represent a region where voxel intensities have been classified as both GM and WM, accounting for this finding in the GM analysis.

Visualisation of this cluster shows it spans the ventral pathway in the temporal lobe. Anatomically, this cluster overlaps with WM in the sagittal stratum, predominantly the ILF and IFOF. These tracts have been proposed to connect with traditional perisylvian language networks via the posterior STG region, and to Broca's area via intralobal fibres (Catani & Mesulam, 2008); however, the precise structural connectivity of the ventral stream is debated (Almairac et al., 2015). In aphasic patients, Xing and colleagues (2017) used tract-based statistics to investigate WM integrity relating to speech comprehension. They found loss of WM integrity throughout the ventral stream, including the IFOF, was related to both word and sentence level comprehension impairments, a finding corroborated in a study of patients with gliomas, where IFOF integrity was related to semantic verbal fluency (Almairac et al., 2015). The correspondence of this cluster to these WM pathways, alongside the nature of the Listen-In task, suggests that change in

this region may be due to stimulation of the speech comprehension network along the ventral pathway. Under this interpretation, repeated exposure to words and sentences would have stimulated this network, instigating neural changes in WM pathways supporting these processes. This stimulation account would explain why density covaried with dose, but not improvements in treated items.

It is important to note, however, that dose was not experimentally manipulated in this study, and that these analyses are post hoc. It is possible that patients lesion damage and changes in brain structure influenced the propensity to do more or less training. In other words, it is not possible to infer whether brain structure influence treatment dose achieved, or whether treatment dose influenced brain structure. Future experimental research would be required to disambiguate between these potential possibilities, with dose as an independent variable.

5.4.3.5 Association between pre-therapy brain structure and treatment-related tissue changes

A further post-hoc analysis (Analysis 5) was carried out in response to neuroplasticity observed in the RH, to explore whether pre-therapy brain structure was associated with magnitude of tissue changes. This was guided by the finding in Chapter 3, which found that localised WM volume in fronto-temporal and subcortical structures predicted improvements in treated item performance. This analysis showed a moderate and positive association between these measures, in that patients with greater WM volume (in cluster regions) at baseline went on to show greater treatment related changes (or lesser atrophy) in GM in the right pSTG. Given that baseline WM correlated with later GM changes, it is possible to infer that WM may be an indicator, or predictor, of potential for plasticity in the ipsilateral hemisphere.

In Chapter 3, a number of possibilities were put forward to account for variability in WM volume, and how this may have functionally related to greater treatment effects. WM variability could relate to premorbid differences in RH speech processing networks; for example, a large minority of individuals have been found to have more bilateral

lateralisation of WM pathways between Broca's and Wernicke's regions, which correlated with better verbal recall (Catani et al., 2007), and so some patients could have been facilitated by existing RH language networks to a greater extent, and demonstrate greater potential for plasticity. Additionally, post-stroke adaptation in these regions prior to Listen-In could have facilitated treatment for some patients, and led to greater potential for further plasticity. In either case, these findings support the notion that greater volume relates to better functioning networks, as patients with greater regional WM volume went on to make greater improvements for treated items, and demonstrated the greatest changes in GM tissue.

The role of the RH in recovery is a matter of debate, and is likely to depend on a number of interacting factors, including phase of recovery, language function, therapy related factors, and lesion related variables (Cocquyt et al., 2017). In the chronic phase of aphasia, the majority of studies have investigated changes in activation patterns in relation to various treatments, with many reporting a facilitatory effect of the RH (Cocquyt et al., 2017). However, in their systematic review, Cocquyt and colleagues (2017) report only two previous studies which found structural changes in response to treatment in the RH. In two of these studies (previously described), facilitatory changes in WM were reported in response to MIT, in the arcuate fasciculous (Schlaug et al., 2009) and an inferior frontal region (Wan et al., 2014). Further evidence for structural plasticity in the RH can be inferred from the cross-sectional study by Xing and colleagues (2016), but also in a longitudinal study by (Hope et al., 2017). The authors followed 28 chronic aphasic patients over the course of one year or more. They found spontaneous changes in word retrieval were positively related to hypertrophy in the anterior temporal lobe. The present study compliments these findings by showing facilitatory changes in GM in the pSTG in the RH, in response to speech comprehension treatment. This supports the notion that change in the RH, in the chronic phase of aphasia, can have a facilitatory effect. It may be that speech processing, by virtue of its proposed bilateral pathways (Hickok & Poeppel, 2007), is better suited to benefit from RH adaptation following stroke. There is evidence to suggest that this may be the case in aphasic patients. Woodhead and colleagues (2017) used MEG to investigate connectivity changes following auditory phonological treatment, and found that in patients with more severe speech comprehension impairments, training

modulated interhemispheric connectivity between left and right STG. In other words, interhemispheric interactions facilitated training effects, suggesting a facilitatory role of the right STG in severe patients. In summary, the present finding provides novel and direct evidence for facilitatory structural adaptation in the RH in the chronic phase of aphasia, and suggests that this RH region could be a potential target for future speech comprehension treatments, such as neural stimulation to enhance activity in the right pSTG.

5.4.3.6 Further considerations

It is possible that pre-morbid individual differences may have contributed to how malleable these regions were in response to training. For example, Mårtensson and colleagues (2012) investigated a group of interpreters on a three-month intensive language training program, learning 300-500 new words each week. Individuals who acquired greater proficiency in the foreign language demonstrated greater increases in cortical thickness in the left mid-posterior STG, as well as changes in volume in the right hippocampus. Although correlational, the authors speculate that the potential of these brain regions to undergo plasticity may explain why some learners show greater aptitude for language learning. In the same way, it may be that patients in Listen-In who had greater potential for plasticity went on to make greater treatment gains. This explanation may account for the linear relation found between treatment gains and structural change, in that degree of plasticity may be distributed across individuals. In Chapter 3, it was suggested that individual differences in learning capacity may account for differences in treatment effects. Building on this, it may be that potential for neuroplasticity shows inter-individual differences across aphasic patients.

5.4.3.7 Strengths of the present study

Previous studies investigating training-induced neuroplasticity have suffered from a number of methodological limitations (Thomas & Baker, 2013). One frequent limitation has been lack of a control group, which makes relating changes to a specific training task problematic. The present study overcame this by using a within-subject design. In this

way, patients acted as their own control, avoiding between-subject variance, and regressing out changes related to time (e.g. due to ageing). Previously, Thomas and colleagues (2009) used a similar within-subject longitudinal design in a simple learning task with healthy individuals. Structural findings were inconsistent and dependent on the processing pipeline. Of note is that the present study used a relatively new serial registration method which has not been frequently employed in intervention studies (Ashburner, 2013). It may be that this method is better suited to identify effects due to superior within subject registration. The plausibility of the regions identified, in relation to the training task, support the interpretation that these are true treatment effects rather than pre-processing related artefacts. The current findings therefore suggest that longitudinal VBM may be a sensitive method for detecting within subject changes over time, and supports the use of this technique in future longitudinal studies.

A further limitation of previous studies has been linking changes to specific tasks. Thomas and Baker (2013) report only seven studies which correlated structural changes with behavioural measures. For example, in Listen-In, it could be that changes represented general exposure to the training task, rather than change on the measure it purported to be testing (I.E. speech comprehension). However, the behavioural variable selected provides strong evidence in favour of the latter, as changes correlated with percentage improvement in speech comprehension, rather than exposure to the task per se. Furthermore, a separate analysis showed significant changes relating to dose, but in a different region, further supporting the specificity of treatment effects in the regions identified.

5.4.3.8 Limitations

The present results demonstrate changes in GM and WM over a treatment period of 12-weeks, over and above changes in a 12-week control period. A further question which arises is the time course of these changes. This was not addressed by the current study, and there have been no systematic studies in humans. Understanding the time course of plasticity is important, as evidence from animals shows structural changes do not occur in a linear fashion with learning. The expansion-partial renormalisation hypothesis suggests

there may be an initial period of expansion ('overshoot'), followed by selective stabilisation and partial renormalisation, wherein the structural changes which remain reflect the learning which has taken place (Lövdén et al., 2013). For example, in a study with three macaque monkeys learning to use a new tool, repeated testing (using VBM) showed GM increases in expected regions which correlated with learning rate; of note was that GM increased the most during the intensive training period, and peaked when performance on the task plateaued (Asamizuya et al., 2009). In Listen-In, the degree of learning which took place correlated with GM and WM changes in regions implicated in speech processing, but the time course of changes, or individual patterns of neuroplasticity, have not been explored. For example, it could be that all patients show the same temporal patterns in neuroplasticity; or, it may be that some patients (e.g. non-responders) show different patterns of plasticity. The answer to this may help us to understand why some patients, who show similar behavioural profiles at baseline, do not respond to particular treatments. In summary, future work could look to systematically investigate the time course of treatment-related structural changes in aphasic patients, and how these may relate to neurobiological models of learning. This information may help to direct schedules of treatment in the future which best capitalise on these mechanisms.

Although the present findings are based on a powerful within-subject design, a limitation is that the control periods were not equal. For Group 1, the within-subject comparison was made with a pre-therapy control period; however, for Group 2, the control period was post-therapy. As previously described, the temporal patterns of experience-dependent neuroplasticity are unclear, which means difference between control periods may have confounded the present findings in some way. For example, in Draganski and colleagues (2004) juggling participants, training induced increases in GM density declined in the three months following cessation of training, and so in Listen-In, patients who received treatment first may have undergone renormalisation of tissue density in their 12-week control period which followed treatment. It is unclear whether this may have influenced the present findings. Of note is that many patients demonstrated declines in GM and WM density in these clusters; one possibility is that these negative values reflect periods of renormalisation, and so indicate regions which have undergone initial

expansion relating to the treatment task. Previously, neuroplasticity has been shown to peak when behaviour is at an asymptotic level, and then decline. If patients show asymptotic performance after, for example, 3 weeks of training, then renormalisation could occur within the 12-week treatment period, and remain hidden. This is not unlikely given that previous speech comprehension treatment studies have found rapid improvements over the first 3-4 sessions, followed by a plateau thereafter (Knollman-Porter et al., 2018). These considerations highlight the difficulty of designing and implementing studies which capture structural changes, particularly when mechanisms in healthy individuals are not yet well understood. Nevertheless, the present findings demonstrate that changes can be identified following 12-weeks of high dose treatment, providing a possible start point for future investigations in aphasic patients.

5.4.3.9 Future directions and clinical implications

Future work into experience dependent plasticity in patients with aphasia could use complimentary neuroimaging techniques to increase our understanding of what brain mechanisms may underlie these changes, particularly given the stroke population is older than typical research populations. For example Draganski and colleagues (2011) used a novel voxel-based quantification (VBQ) technique, which used a wide number of contrast parameters during MRI which reveal the physical properties of water, and thus tissue composition. In their study of normal ageing, volume changes found using VBM were supported by VBQ results which showed specific changes in four of these contrasts, providing additional information about underlying tissue properties. In the future, 7T scanners may also be useful for detecting treatment related changes, as the higher magnetic field produces more detailed and higher quality images, which may in turn lead to better detection of subtle changes in tissue.

One possible clinical direction for treatments like Listen-In may be in promoting maintenance of tissue in the residual language network, by stimulating language pathways. For example, Hope and colleagues (2017) found improvements as well as declines in their chronic aphasic patients' language skills over one year or more, which were systematic and predictable given associated structural changes in the RH. These

findings are in line with those reported here, which demonstrate both directions of structural change in relation to treatment outcomes. One possibility, given the present finding, is that stimulation of language networks could help to maintain residual function. Given these preliminary findings, this role is speculative; however, it offers a future possible avenue to explore given the new era of digital based interventions which would allow for frequent self-administered use. In this way, some treatments could aim to halt gradual declines in language skills, rather than improve them per se.

The present finding provides direct evidence for a facilitatory role of the RH during treatment. The particular region identified could therefore be a potential target for future speech comprehension treatments. For example, non-invasive neural stimulation could enhance activity in the right pSTG, and possibly augment behavioural effects. Evidence has increasingly demonstrated that stimulation, such as transcranial direct current stimulation (tDCS), can have positive effects on aphasia rehabilitation, but the majority of studies have focused on speech output (Biou et al., 2019). The present finding provides a rationale for targeting this region in future stimulation studies for speech comprehension.

6 General Discussion

6.1 Summary of key results

In this thesis I have reported the development of a speech comprehension therapy app, co-designed with persons with post-stroke aphasia ('Listen-In'). I then reported investigations into the efficacy of Listen-In for treating speech comprehension impairments in chronic aphasic patients. Lastly, I carried out investigations into brain and behavioural factors which related to treatment outcomes.

The main aims of this thesis were as follows:

- The first aim, in Chapter 1, was to develop a tablet-based therapy application targeting speech comprehension impairments, at the single word level, for persons with post-stroke aphasia. A novel aspect of this approach was the implementation of gamification. This development stage involved an iterative co-design process with persons with aphasia, and the software development team.
- In Chapter 2, my aim was to test the efficacy of Listen-In in group of chronic aphasic patients with speech comprehension impairments. A second aim was to investigate whether baseline measures could explain variability in response to treatment.
- My next aim, in Chapter 3, was to investigate whether structural integrity of pre-treatment brain tissue could predict response to Listen-In treatment.
- My final aim, in Chapter 4, was to investigate therapy driven structural neuroplasticity.

In this thesis, I developed and tested a novel tablet-based treatment designed to improve comprehension of spoken words in patients with chronic aphasia. The therapeutic component consisted of spoken word/phase/sentence-to-picture matching challenges, which were individually curated, to form over 3000 unique therapy items. The iterative co-design process with patients with aphasia, and software developers, resulted in a

series of focus groups, and multiple iterations of the prototype, over the period of approximately one year. As described in **Chapter 1**, many barriers to usability and enjoyability were identified and rectified, following feedback and observations from group members. Importantly, many of these barriers were unanticipated, demonstrating that the ideas of well-intentioned researchers, clinicians and software developers do not always meet the needs or preferences of persons with aphasia. This finding likely extends to other patient populations. Involving end users in the design of therapies is therefore vital, as one group member commented: *"Isn't it obvious? You have to involve us in the design phase, because we will be the ones who end up using it"*. A key challenge of gamification was developing a game which met the needs of a range of individuals with different preferences and aphasia severities. A related, and unexpected, challenge was the differing views of individuals to gamification; some patients enjoyed this aspect, but some patients failed to understand the point of the game. Ultimately, it was not possible to meet all users' individual preferences. However, this process resulted in an app which, crucially, was able to used independently by all members of the group, regardless of severity of language impairment.

Despite the wealth of apps and computer programs available for individuals with aphasia (https://www.aphasiasoftwarefinder.org/), few have reported co-designing treatments for persons with aphasia. A research group at City University have recently pioneered this approach, involving persons with aphasia in the co-design of a number of digital applications, demonstrating the feasibility but also challenges of this method (Galliers et al., 2017; Marshall et al., 2013). The present study builds on this work by demonstrating the benefits of co-design in the development of a treatment app with gamification, a feature novel to aphasia therapies. The use of co-design has undoubtedly contributed to a superior final product which better meets the requirements of this population. Ultimately, this process has highlighted that patients with even severe language impairments can and should be involved in the development stage of these types of products. Although challenges were encountered, the previous co-design studies, and now Chapter 1 of this thesis, demonstrate that careful planning and flexibility can support the involvement of a wide range of persons with aphasia in research.

In **Chapter 2** I tested the efficacy of Listen-In in a randomised cross-over repeated measures trial with 35 individuals with chronic post-stroke aphasia. All individuals had impairments in understanding spoken words and sentences, and the aim of treatment was to improve comprehension of single words. The treatment was well tolerated by patients who self-administered a large mean dose of 85 hours, the largest dose of treatment which, to my knowledge, has been reported for speech comprehension therapy. The results showed that patients made large and significant improvements in their comprehension of words which were trained within the Listen-In app. Previously these types of impairments have been regarded by some as resistant to treatment. The present findings therefore offer key contributions to the field by showing that: motivated patients are willing and able to complete high dose treatments if given the opportunity; significant gains in spoken word comprehension can be made even years after stroke; therapy gains are maintained after treatment stops (up to 24 weeks). Indeed, the participant who made the largest gains had been living with aphasia for over twenty years.

The use of technology as a medium to deliver treatment has made this magnitude of dose possible, demonstrated by the fact that the average patient completed over twenty thousand individual challenges over eighty-five hours of treatment. These findings support computerised delivery of high dose aphasia treatments (Bhogal et al., 2003; Brady et al., 2016). Currently, dose of treatment is a key issue within stroke rehabilitation. This was recently demonstrated by the Stroke Association's Amazing Brains: Road to Recovery event (2019), where Professor Nick Ward emphasised the important of high dose treatments to build on principles of neuroplasticity, and filmmaker, Fiona Lloyd-Davies, described a personal account of motor improvements in a patient following high dose rehabilitation. Unfortunately, there are currently few studies which have investigated these high doses, and studies are needed to explore the full potential of these regimes if they are to be implemented into current practice, and benefit patients. In this respect, the present findings offer new evidence, within the speech comprehension domain, for high dose treatment. Listen-In was repetitious, intensive, and specific, all characteristics which support neuroplasticity (Kleim & Jones, 2008), and in line with this, patients evidenced changes both at the behavioural and neural level (Chapter 4). Future impairment-based treatment studies must consider these principles to ensure

conclusions regarding efficacy of treatments are based on treatments which have provided a sufficient dose to instigate change.

A caveat to Listen-In improvements was that they did not generalise to untrained words, or to improvements on any other outcome measure. This demonstrates that therapy supported learning through item-specific mechanisms. Item specific treatment effects are ubiquitous amongst other therapy approaches which train single items, such as naming and gesture treatments (Marshall et al., 2013; Webster et al., 2015). One criticism of these findings is the functional relevance of these gains to real life communication contexts. It is also unclear from these results whether patients could generalise comprehension of items they did learn, to novel contexts (however, see further directions). This raises two avenues for further developments and investigations: firstly, to capitalise on item specific treatment effects by incorporating therapy items which are meaningful for individual patients; and investigating whether manipulations of treatment parameters might support different kinds of generalisation. The latter has been neglected in speech comprehension in aphasia research, and offers a potential new avenue of systematic research.

A key characteristic of these behavioural findings was considerable variability in magnitude of treatment effects; not all patients responded to the same extent, and some didn't respond at all. An important consideration of these types of treatments is investigating what type of treatment works, and for whom, particularly given the time and energy expense for high dose treatments which patients are expected to undertake, often based on clinical recommendations. Clinicians therefore need access to information which can help them to make evidence-based judgements on when best to prescribe particular types of treatments. Unfortunately, no baseline factors have been consistently identified in the literature. Accordingly, I found that severity of speech comprehension impairment, and time spent on treatment, bore no relation with treatment outcomes, and although a combination of behavioural factors were identified, these only accounted for a small amount of variability in outcomes. However, the pattern of auditory and phonological measures which were included as the top predictors in the model suggests a link between these skills and response to treatment; this needs further investigation. In light of this, a

recent line of enquiry being undertaken by the PLORAS project is understanding how brain structure may aid in predicting outcomes (Price et al., 2010).

In **Chapter 3**, I therefore extended my previous analysis by investigating whether pre-treatment tissue integrity could account for variability in treatment outcomes. Previously, a small number of studies have suggested that integrity of peri-lesional regions in the LH can impact treatment outcomes, but none have investigated this for speech comprehension (Aguilar et al., 2018; Fridriksson, 2010; Meinzer et al., 2010). My results were unexpected. I found no areas in the LH which related to treatment outcomes. Instead, I found mainly white matter regions, distributed throughout the RH, which were important for treatment outcomes. Of note was that these regions corresponded to classic language homologues in the temporal and frontal lobes, as well as regions which corresponded to learning and memory networks. Overall, these appeared to represent multifunctional networks. I also found that patients with greater volume in these regions tended to make greater improvements, leading to the conclusion that greater volume equals greater function. Given the location of these structural findings, in key language, learning and memory regions, I speculated that these white matter regions reflect multi-functional networks which likely support treatment through a range of language and cognitive functions.

These findings highlight a distinction which has not commonly been investigated in aphasia therapy: the level of language ability or impairment, versus the ability to learn. In the present study, RH regions correlated with a behavioural measure which ultimately reflects the ability to learn or re-learn spoken words, rather than speech comprehension per se. In this way, RH networks appear to play an important role in the learning process during language treatment. Whether this is unique to speech comprehension therapy, perhaps due to greater bilateral organisation of speech processing, or relevant to learning mechanisms in general, needs further investigation.

This RH finding was underpinned by the presence of variability in WM volume amongst patients. A key question which emerged was whether variability in volume of RH WM networks related to pre-morbid individual differences, or post-stroke neural

reorganisation. One hypothesis is that variability relates to premorbid individual differences. An alternative hypothesis is that variability reflects post-stroke neural reorganisation.

In **Chapter 4**, I directly investigated neural reorganisation by investigating structural neuroplasticity in response to Listen-In treatment. My key findings showed that: improvements in speech comprehension were related to localised tissue changes in bilateral speech processing regions in the temporal lobes; and exposure to the task was related to tissue changes in white matter underlying the ventral speech processing pathway. Of note were these patients were years post stroke, showing that patients can make significant behavioural improvements, supported by neural plasticity, if given the opportunity to practice a particular skill in sufficient quantity.

The locations of tissue change are in key parts of the speech processing network in healthy individuals, and directly relate to the treatment task, strongly supporting the validity of these findings. To my knowledge, this longitudinal VBM method has not been used in aphasia research, and rarely reported in other patient populations (Ashburner, 2013). These findings therefore offer novel evidence which show this approach is sensitive to tissue changes in this population, and suggests this method may be suitable to investigate adaptation in further patient treatment studies.

The role of the right hemisphere is aphasia recovery is currently a matter of much debate. These findings contribute to this debate by showing that patients engage both hemispheres in a facilitatory manner during speech comprehension training, supporting an adaptive role of the RH. It may be that this type of treatment is particularly amenable to RH compensation due to the posited bilateral organisation of spoken word recognition (Hickok & Poeppel, 2000, 2004, 2007). In this way, individual pre-morbid variation in these RH networks (Catani et al., 2007) may have put some patients at an advantage for treatment. The findings from Chapter 3 support this account, as they show that greater volume of RH WM was associated with better behavioural outcomes; Chapter 4 extends this by showing that greater volume of RH WM is also associated with greater therapy-driven plasticity in the RH. Furthermore, a common region of neuroplasticity in the RH

lends support to the notion that this structure had a common function prior to treatment, which patients were engaging during Listen-In. In summary, the findings from Chapters 3 and 4 suggest that it may be pre-morbid individual differences in network structure and/or function in the RH which contribute to variability in treatment outcomes.

There is a growing interest in investigating the function of RH networks in response to treatment, and evidence suggest treatments may modulate network activity in the RH (Cocquyt et al., 2017). The present findings contribute to this literature by providing convergent evidence which shows that functional changes in the RH likely lead to longer term changes in structure. Future investigations may look at reconciling the location of plasticity observed here, with functional changes before and after this type of treatment, to elucidate the relation between structure and function in speech comprehension treatment.

6.2 Future Directions

6.2.1 Development of Listen-In application

A number of challenges with gaming and therapy features were encountered during the trial, from the research perspective in terms of the therapy data, but also from the patient perspective. Following on from these, re-development is now underway, and has included a number of key changes, summarised below. Following re-development, the app will be made available to the public on the app store, for patients to independently download and use.

Changes relating to gamification

- In line with findings from the focus groups in Chapter 1, patients' responses to the gaming component were mixed. For some, it was an enjoyable break from the therapy challenges, but for others, it became a barrier to therapy and led to considerable frustration. An option to skip the game component after each therapy block will be implemented to reduce frustration for those who do not want to engage with the game.

- A frequent response was that some patients found the theme childlike. Although it was not possible to accommodate all individual preferences, the frequency of this response suggests that this could be a barrier to future uptake by patients. The theme of the app has therefore been re-developed to a less child-like character (Figure 6-1).

Figure 6-1 Old and new character design for Listen-In

Changes relating to progression through therapy items

- There was considerable variability in how therapy items were presented, both within and across patients, in terms of number of exposures, and timing of exposures, as shown in Chapter 2. In response to this, a new ladder of progression has been developed which will fix the number of times a lexical item is trained. Timing of item exposure will be pseudorandomised to ensure even presentation throughout a treatment cycle.
- The adaptive algorithm automated progression up and down challenges based on pre-determined difficulty. Difficulty will now be implemented as a user option, where players can choose to practice 'easy' or 'hard' challenges for groups of

therapy items. This will ensure patients have some control over the level of difficulty.

Changes relating to measuring outcomes

- Trained items tested in the ACT outcome measure were always tested in the same linguistic context as those presented in treatment. One criticism is that patients may be learning words in a context specific fashion, rather than generalising comprehension of words across linguistic contexts. This is an important consideration, as if patients are learning words in specific constructions, then relevance to everyday communication is questionable. In the next version of Listen-In, items tested in the ACT will not be trained during treatment. This will ensure that improvements, if seen, represent comprehension of the word across different linguistic contexts.

6.2.2 Enhancing treatment effects

In Listen-In, an average patient completed around 22000 individual challenges over 85 hours. However, it is unclear how the content and delivery of challenges contributed to treatment outcomes. In this respect Listen-In could be viewed as a 'black box', as it is not clear what combination of treatment variables contribute to maximal gains. It is possible that manipulating these could enhance the size of treatment effects. Examples of these variables are total number of therapy challenges, total number of exposures per item, number of challenges per item, number of exposures required during the learning and then maintenance phases, and pattern of exposures over time. Different combinations could be tested in the future to investigate which is most conducive to learning. For example, one option could be to train matched sets of items in different ways and compare outcomes, within an individual patient. Alternatively, groups of patients could be randomised to receive different versions.

A further investigation is currently underway to assess the contribution of exposing different linguistic constructions for target words, on outcomes. This analysis was based on frequency data for exposures of different types of constructions during treatment, for

each patient. Hierarchical logistic regression identified a winning model which included contributions of both the exact linguistic construction, as well as different forms of linguistic constructions, to treatment outcomes. In other words, training the word in different contexts supported positive outcomes.

6.2.3 Investigating patterns of performance

The present study set out to test whether Listen-In could improve spoken word comprehension at a group level. As such, a detailed investigation into individual response to treatment has not been conducted in the present thesis. It is possible that this type of investigation may highlight why some individuals responded to treatment and why some didn't, which was not apparent in the group level analyses. For example, regression models, as conducted in Chapter 2, rely on data with a linear association, but it may be that subgroups of patients show particular patterns of baseline performance and treatment outcomes which don't conform to a linear association. This may be particularly relevant considering the heterogeneity of the present cohort in terms of aphasia profiles. A case series analyses could be used to investigate these questions.

A fine-grained investigation of selected behavioural assessments may also provide additional insight. For example, the co-primary outcome measure (ACT) used a single accuracy score to represent performance. However, reaction times, and number of repetitions, were also captured. A score which incorporates all aspects of these parameters could provide more fine-grained thresholds of performance rather than accuracy alone. For example, an individual may require less repetitions which could be viewed as an improvement, even if overall accuracy did not improve. In real life terms, requiring less repetitions would also be beneficial. Two further auditory processing measures also included additional parameters. The Phoneme Discrimination Test (PDT) was developed with different levels of phoneme similarity in terms of voice, placing and manner, and in an adaptive version, thresholds were found to significantly relate to speech comprehension ability (Robson, Keidel, et al., 2012). In the environmental sounds test (ENVASA), a number of conditions were also included, such as congruency and signal-

to-noise ratio, which may provide further insight into level of non-verbal auditory processing ability.

6.2.4 Predicting response to Listen-In treatment

In Chapter 2, a large combination of baseline factors explained some of the variance in treatment outcomes; however, this was small. In Chapter 3, I predicted that structural integrity of peri-lesional regions in the LH would be associated with treatment outcomes. Instead, greater volume of WM networks in the RH related to treatment outcomes. However, it is possible that VBM was insensitive to identifying LH regions due to the often large and overlapping lesions in the current cohort, which means some regions may have little or no tissue variability. Aguilar and colleagues (2018) used a method which parcellated patients brains into a large number of regions, and measured the proportion damage to each region for each patient. Using automated linear modelling, as used in Chapter 2, the authors found that a combined model of behavioural, demographic, and lesion damage in the LH, best explained response to a reading treatment in patients with central alexia. A future avenue of investigation for the current data would be to replicate these methods. In combination with the predictive role of the RH, these may provide convergent evidence which further elucidates the roles of both hemispheres in speech comprehension recovery.

In Chapter 2, ALM was able to explain a small proportion of variance in treatment outcomes. However, this model explained data from the current cohort, but did not make predictions regarding new patients. A further option, also reported by Aguilar and colleagues (2018) is to conduct an out of sample analysis. This method removes patients from the cohort, and using a model based on the remaining patients, attempts to predict response to therapy for this excluded patient (an out of sample analysis). This has significant advantages over explanatory models which may suffer from overfitting, and lead to poor predictions for future patients. The current sample of 35 may be too small for this type of analysis; however, this may provide an interesting starting point for future studies attempting to predict treatment outcomes.

A key question which emerged in Chapter 3, was whether variability in volume of RH WM networks related to pre-morbid individual differences, or post-stroke adaptation. My hypothesis is that patients show the same volume as controls, reflecting pre-morbid individual differences in WM networks. However, an alternative hypothesis is that volume of WM clusters in the RH is greater than in age-matched controls. The rationale for this hypothesis is that patients may undergo considerable neuroplasticity in RH networks following stroke, and that patients who show greater reorganisation may be better able to engage these networks to support re-learning during treatment. For example, Xing and colleagues (2016) found that volume of GM regions in the RH, which correlated with language measures, was greater in patients than in controls, suggesting reorganisation had taken place. To investigate these hypothesis, future work could compare patients' RH tissue volume with healthy and non-aphasic stroke patients, matched for scanner type, and demographic variables. Understanding the role of the RH in recovery is important so that in the future, treatments may be able to build on these reorganisation patterns.

6.3 Limitations

The present thesis investigated therapy which targeted comprehension at the level of the single word. As such, the literature reviewed, and interpretation of findings, has mainly involved evidence relating single word processing. However, in the therapy task and outcome measure words were also embedded within phrases and sentences to provide greater variability in linguistic context, more akin to typical speech, and also to manipulate level of difficulty. As such, these analyses were not a pure investigation of single word processing. For example, patients with greater sentence processing ability may have been able to use this to support their word comprehension. Words which were tested within longer sentence may have been more resistant to comprehension and/or treatment as they would have entailed greater processing demands such as auditory short-term memory. Linguistic context may have made some words easier to comprehend, for example, if the target word was more congruent with the verb (e.g. "she throws the ball" versus "she throws the bell"). It is also not clear how sentence to picture mapping tasks support word form-meaning mappings, given the extra step of identifying which part of the sentence matches the picture. Given these extra considerations, improvements in

comprehension, as found here, may actually appear more robust, as many other factors may have interfered to make comprehension and learning more difficult. Future investigations are needed to examine how providing linguistic contexts impacts comprehension and learning of single words, and whether this supports different types of generalisation.

A further limitation of the present study is that the impact of therapy on patients' functional communication in everyday life was not fully explored. The measure which was developed to investigate this question showed considerable within and across subject variability, suggesting poor reliability. Therefore it was unclear whether improvements translated to everyday situations. Exit-interviews provided some insight into this impact, however, these relied on patients having an interested relative or carer and so was not possible for all patients. These types of informal feedback methods are also subject to the typical biases inherent in qualitative research, such as participants reporting positive outcomes to please the researcher. New ways of measuring the impact of speech comprehension treatment are needed; however, given the hidden nature of speech comprehension, this is a challenge. One way may be to explore carer and relative observations in a more robust and systematic manner, such as with semi-structured interviews with researchers unconnected to the study.

7 References

Abel, S., Weiller, C., Huber, W., Willmes, K., & Specht, K. (2015). Therapy-induced brain reorganization patterns in aphasia. *Brain*, *138*(4), 1097–1112.

Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. In *Journal of Neurolinguistics* (Vol. 20).

Aguilar, O. M., Kerry, S. J., Ong, Y.-H., Callaghan, M. F., Crinion, J., Woodhead, Z. V. J., … Hope, T. M. H. (2018). Lesion-site-dependent responses to therapy after aphasic stroke. *Journal of Neurology, Neurosurgery & Psychiatry*, *89*(12), 1352–1354.

Almairac, F., Herbet, G., Moritz-Gasser, S., de Champfleur, N. M., & Duffau, H. (2015). The left inferior fronto-occipital fasciculus subserves language semantics: a multilevel lesion study. *Brain Structure and Function*, *220*(4), 1983–1995.

Altman, D. G., & Bland, J. M. (2005). Treatment allocation by minimisation. *BMJ (Clinical Research Ed.)*, *330*(7495), 843.

Anguera, J. A., Boccanfuso, J., Rintoul, J. L., Al-Hashimi, O., Faraji, F., Janowich, J., … Gazzaley, A. (2013). Video game training enhances cognitive control in older adults. *Nature*, *501*(7465), 97–101.

Archibald, L. M. D., Orange, J. B., & Jamieson, D. J. (2009). Implementation of computer-based language therapy in aphasia. *Therapeutic Advances in Neurological Disorders*, *2*, 299–311.

Asamizuya, T., Quallo, M. M., Iriki, A., Cheng, K., Lemon, R. N., Ueno, K., & Price, C. J. (2009). Gray and white matter changes associated with tool-use learning in macaque monkeys. *Proceedings of the National Academy of Sciences*, *106*(43), 18379–18384.

Ashburner, J. (2013). Symmetric diffeomorphic modeling of longitudinal structural MRI. *Frontiers in Neuroscience*, *6*, 197.

Ashburner, J., & Friston, K. J. (2000). Voxel-based morphometry - The methods. *NeuroImage*, *11*(6 I), 805–821.

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage*, *26*(3), 839–851.

Baker, E., Blumstein, S. E., & Goodglass, H. (1981). Interaction between phonological and semantic factors in auditory comprehension. *Neuropsychologia*, *19*(1), 1–15.

Basso, A. (2003). *Aphasia and its Therapy*. Oxford University Press: New York.

Bastiaanse, R., Nijboer, S., & Taconis, M. (1993). *The Auditory Language Comprehension Programme: A description and case study*. *28*(4), 415–433.

Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., & Dronkers, N. F. (2003). Voxel-based lesion–symptom mapping. *Nature Neuroscience*, *6*(5), 448–450.

Behrmann, M., & Lieberthal, T. (1989). Category-specific treatment of a lexical-semantic deficit: a single case study of global aphasia. *The British Journal of Disorders of Communication*, *24*(3), 281–299.

Best, W., Greenwood, A., Grassly, J., Herbert, R., Hickin, J., & Howard, D. (2013). Aphasia rehabilitation: Does generalisation from anomia therapy occur and is it predictable? A case series study. *Cortex*, *49*(9), 2345–2357.

Best, W., & Howard, D. (1994). *Word sound deafness resolved? 8*(3), 223–256.

Bezzola, L., Merillat, S., Gaser, C., & Jancke, L. (2011). Training-Induced Neural Plasticity in Golf Novices. *Journal of Neuroscience*, *31*(35), 12444–12448.

Bhogal, S. K., Teasell, R., & Speechley, M. (2003). Intensity of Aphasia Therapy, Impact on Recovery. *Stroke*, *34*(4), 987–993.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Springer, J. A., Kaufman, J. N., & Possing, E. T. (2000). Human Temporal Lobe Activation by Speech and Nonspeech Sounds. *Cerebral Cortex*, *10*, 512–528.

Binney, R. J., Pankov, A., Marx, G., He, X., McKenna, F., Staffaroni, A. M., … Rosen, H. J. (2017). Data-driven regions of interest for longitudinal change in three variants of frontotemporal lobar degeneration. *Brain and Behavior*, *7*(4), 1–11.

Biou, E., Cassoudesalle, H., Cogne, M., Sibon, I., Gabory, I. De, Dehail, P., … Glize, B. (2019). Transcranial direct current stimulation in post-stroke aphasia rehabilitation: a systematic review. *Annals of Physical and Rehabilitation Medicine*, (2018).

Blumstein, S. E., Baker, E., & Goodglass, H. (1977). Phonological factors in auditory comprehension in aphasia. *Neuropsychologia*, *15*(1), 19–30.

Boekel, W., Wagenmakers, E. J., Belay, L., Verhagen, J., Brown, S., & Forstmann, B. U. (2015). A purely confirmatory replication study of structural brain-behavior correlations. *Cortex*, *66*, 115–133.

Bowen, A., James, M., & Young, G. (2016). *National clinical guideline for stroke*.

Brady, M. C., Kelly, H., Godwin, J., Enderby, P., & Campbell, P. (2016). Speech and language therapy for aphasia following stroke. *Cochrane Database of Systematic Reviews*, (6), 4–7.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

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

Breier, J. I., Juranek, J., Maher, L. M., Schmadeke, S., Men, D., & Papanicolaou, A. C. (2009). *Behavioral and Neurophysiologic Response to Therapy for Chronic Aphasia*. *90*(12), 2026–

2033.

Breitenstein, C., Grewe, T., Flöel, A., Ziegler, W., Springer, L., Martus, P., … Bamborschke, S. (2017). Intensive speech and language therapy in patients with chronic aphasia after stroke: a randomised, open-label, blinded-endpoint, controlled trial in a health-care setting. *The Lancet*, *389*(10078), 1528–1538.

Breitenstein, C., Jansen, A., Deppe, M., Foerster, A. F., Sommer, J., Wolbers, T., & Knecht, S. (2005). Hippocampus activity differentiates good from poor learners of a novel lexicon. *NeuroImage*, *25*(3), 958–968.

Broca, P. (1861a). Nouvelle observation d'aphémie produite par une lésion de la troisième circonvolution frontale. *Bulletins de La Société d'anatomie (Paris)*, *6*(2e serie), 398–407.

Broca, P. (1861b). Perte de la parole: ramollissement chronique et destruction partielle du lobe anterieur gauche du cerveau. *Bulletins de La Societe d'anthropologie, 2*, *2*((1re serie),), 235–238.

Broca, P. (1864). Sur les mots aphemie, aphasie et aphrasie. *Lettre a M. Le Professeur Trousseau, Gazette Des Hopitaux*, *23*.

Brownsett, S. L. E., Warren, J. E., Geranmayeh, F., Woodhead, Z., Leech, R., & Wise, R. J. S. (2014). Cognitive control and its impact on recovery from aphasic stroke. *Brain*, *137*(1), 242–254.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911.

Bubb, E. J., Kinnavane, L., & Aggleton, J. P. (2017). Hippocampal–diencephalic–cingulate networks for memory and emotion: An anatomical guide. *Brain and Neuroscience Advances*, *1*, 239821281772344.

Buchel, C., Gaser, C., Boyke, J., Driemeyer, J., & May, A. (2008). Training-Induced Brain Structure Changes in the Elderly. *Journal of Neuroscience*, *28*(28), 7031–7035.

Buchsbaum, B. R., Baldo, J., Okada, K., Berman, K. F., Dronkers, N., D'Esposito, M., & Hickok, G. (2011). Conduction aphasia, sensory-motor integration, and phonological short-term memory - An aggregate analysis of lesion and fMRI data. *Brain and Language*, *119*(3), 119–128.

Burton, E. J., Kenny, R. A., O'Brien, J., Stephens, S., Bradbury, M., Rowan, E., … Ballard, C. G. (2004). White matter hyperintensities are associated with impairment of memory, attention, and global cognitive performance in older stroke patients. *Stroke*, *35*(6), 1270–1275.

Butler, R. A., Lambon Ralph, M. A., & Woollams, A. M. (2014). Capturing multidimensionality in

stroke aphasia: mapping principal behavioural components to neural structures. *Brain*, *137*(12), 3248–3266.

Callan, D. E., Kawato, M., Parsons, L., & Turner, R. (2007). Speech and song: The role of the cerebellum. *Cerebellum*, *6*(4), 321–327.

Caplan, D., Waters, G., DeDe, G., Michaud, J., & Reddy, A. (2007). A study of syntactic processing in aphasia I: Behavioral (psycholinguistic) aspects. *Brain and Language*, *101*(2), 103–150.

Carreiras, M., Seghier, M. L., Baquero, S., Estévez, A., Lozano, A., Devlin, J. T., & Price, C. J. (2009). An anatomical signature for literacy. *Nature*, *461*(7266), 983–986.

Catani, M., Allin, Matthew, P. G., Husain, M., Pugliese, L., Mesulam, M. M., Murray, R. M., & Jones, D. K. (2007). Symmetries in human brain language pathways correlate with verbal recall. *Proceedings of the National Academy of Sciences*, *104*(43), 17163–17168.

Catani, M., Jones, D. K., & Ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Annals of Neurology*, *57*(1), 8–16.

Catani, M., & Mesulam, M. (2008). The arcuate fasciculus and the disconnection theme in language and aphasia: History and current state. *Cortex*, *44*(8), 953–961.

Cattell, R. (1940). A culture free intelligence test 1. *Annual Review of Psychology*, (3), 161–179.

Chein, J. M., & Schneider, W. (2005). Neuroimaging studies of practice-related change: fMRI and meta-analytic evidence of a domain-general control network for learning. *Cognitive Brain Research*, *25*(3), 607–623.

Cherney, L. R., Patterson, J. P., & Raymer, A. M. (2011). Intensity of aphasia therapy: Evidence and efficacy. *Current Neurology and Neuroscience Reports*, *11*(6), 560–569.

Chwat, S., & Gurland, G. B. (1981). *Comparative family perspectives on aphasia: diagnostic, treatment, and counseling implications* (pp. 212–225). pp. 212–225. BRK Publishers.

Clark, I. A., Kim, M., & Maguire, E. A. (2018). Verbal Paired Associates and the Hippocampus: The Role of Scenes. *Journal of Cognitive Neuroscience*, *30*(12), 1821–1845.

Cocquyt, E.-M., De Ley, L., Santens, P., Van Borsel, J., & De Letter, M. (2017). The role of the right hemisphere in the recovery of stroke-related aphasia: A systematic review. *Journal of Neurolinguistics*, *44*, 68–90.

Code, C., & Heron, C. (2003). Services for aphasia, other acquired adult neurogenic communication and swallowing disorders in the United Kingdom, 2000. *Disability and Rehabilitation*, *25*(21), 1231–1237.

Code, C., & Heron, C. (2009). Services for aphasia, other acquired adult neurogenic communication and swallowing disorders in the United Kingdom, 2000. *Disabil Rehabil*, *25*(21), 1231–1237.

Coltheart, M., Patterson, K., & Marshall, J. C. (1987). *Deep Dyslexia*. New York: Routledge & Kegan Paul Inc.

Crosson, B., McGregor, K., Gopinath, K. S., Conway, T. W., Benjamin, M., Chang, Y. L., … White, K. D. (2007). Functional MRI of language in aphasia: A review of the literature and the methodological challenges. *Neuropsychology Review*, *17*(2), 157–177.

Cruice, M., Worrall, L., Hickson, L., Hirsch, F., & Holland, A. (2013). *Quality of life for people with aphasia: performance on and usability of quality of life assessments*. *5*(2), 85–91.

De Schutter, B. (2011). Never too old to play: The appeal of digital games to an older audience. *Games and Culture*, *6*(2), 155–170.

DeDe, G. (2012). Effects of Word Frequency and Modality on Sentence Comprehension Impairments in People With Aphasia. *American Journal of Speech-Language Pathology*, *21*(2), S103–S114.

DeWitt, I., & Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences*, *109*(8), E505–E514.

Dignam, J., Copland, D., McKinnon, E., Burfein, P., O'Brien, K., Farrell, A., & Rodriguez, A. D. (2015). Intensive Versus Distributed Aphasia Therapy. *Stroke*, STROKEAHA.115.009522.

Dignam, J., Copland, D., Rawlings, A., Kate, O., Burfein, P., & Rodriguez, A. D. (2016). *The relationship between novel word learning and anomia treatment success in adults with chronic aphasia. 81*, 186–197.

Dignam, J., Rodriguez, A. D., & Copland, D. A. (2016). Evidence for Intensive Aphasia Therapy: Consideration of Theories From Neuroscience and Cognitive Psychology. *PM & R : The Journal of Injury, Function, and Rehabilitation*, *8*(3), 254–267.

Dole, M., Meunier, F., & Hoen, M. (2013). Gray and White Matter Distribution in Dyslexia: A VBM Study of Superior Temporal Gyrus Asymmetry. *PLoS ONE*, *8*(10), 1–14.

Dorze, G. Le, & Brassard, C. (1995). A description of the consequences of aphasia on aphasic persons and their relatives and friends, based on the WHO model of chronic diseases. *Aphasiology*, *9*(3), 239–255.

Draganski, B., Ashburner, J., Hutton, C., Kherif, F., Frackowiak, R. S. J., Helms, G., & Weiskopf, N. (2011). Regional specificity of MRI contrast parameter changes in normal ageing revealed by voxel-based quantification (VBQ). *NeuroImage*, *55*(4), 1423–1434.

Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Neuroplasticity: Changes in grey matter induced by training. *Nature*, *427*(6972), 311–312.

Draganski, B., Kherif, F., Klöppel, S., Cook, P., Alexander, D., Parker, G. J. M., … Frackowiak, R. S. J. (2008). *Evidence for Segregated and Integrative Connectivity Patterns in the Human Basal*

Ganglia. *28*(28), 7143–7152.

Dronkers, N. F., Wilkins, D. P., Valin, R. D., Redfern, B. B., & Jaeger, J. J. (2004). *Lesion analysis of the brain areas involved in language comprehension*. *92*(1–2), 145–177.

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *British Picture Vocabulary Scale (2nd ed.)*. NFER-Nelson.

Ellis, A., Miller, D., & Sin, G. (1983). Wernlcke's aphasia and normal language processing: A case study in cognitive neuropsychology*. *Cognition*, *15*, 111–144.

Ellis, A., & Young, A. (1988). *Human Cognitive Neuropsychology*. Erlbaum: Hove.

Eriksson, S. H., Free, S. L., Thom, M., Symms, M. R., Martinian, L., Duncan, J. S., & Sisodiya, S. M. (2009). Quantitative grey matter histological measures do not correlate with grey matter probability values from in vivo MRI in the temporal lobe. *Journal of Neuroscience Methods*, *181*(1), 111–118.

Evans, S., Kyong, J. S., Rosen, S., Golestani, N., Warren, J. E., McGettigan, C., … Scott, S. K. (2014). The Pathways for Intelligible Speech: Multivariate and Univariate Perspectives. *Cerebral Cortex*, *24*(9), 2350–2361.

Fedorenko, E. (2014). The role of domain-general cognitive control in language comprehension. *Frontiers in Psychology*, *5*, 335.

Fedorenko, E., Duncan, J., & Kanwisher, N. (2012). Language-selective and domain-general regions lie side by side within Broca's area. *Current Biology*, *22*(21), 2059–2062.

Field, A. (2013). *Discovering Statistics using IBM SPSS statistics; and sec and drugs and rock "n" roll (4th ed.)*. Sage Publications: Portland.

Fiez, J. A., Damasio, H., & Grabowski, T. J. (2000). Lesion segmentation and manual warping to a reference brain: Intra- and interobserver reliability. *Human Brain Mapping*, *9*(4), 192–211.

Fillingham, J. K., Hodgson, C., Sage, K., & Lambon Ralph, M. A. (2003). The application of errorless learning to aphasic disorders: A review of theory and practice. *Neuropsychological Rehabilitation*, *13*(3), 337–363.

Fillingham, J. K., Sage, K., & Lambon Ralph, M. a. (2006). The treatment of anomia using errorless learning. *Neuropsychological Rehabilitation*, *16*(2), 129–154.

Flandin, G., & Friston, K. J. (2015). Topological Inference. In *Brain Mapping: An Encyclopedic Reference* (pp. 495–500). Cambridge: Massachusetts: Academic Press.

Flöel, A., de Vries, M. H., Scholz, J., Breitenstein, C., & Johansen-Berg, H. (2009). White matter integrity in the vicinity of Broca's area predicts grammar learning success. *NeuroImage*, *47*(4), 1974–1981.

Francis, D. R., Jane Riddoch, M., & Humphreys, G. W. (2001). Cognitive rehabilitation of word

meaning deafness. *Aphasiology*, *15*(8), 749–766.

Franklin, S. (1989). Dissociations in auditory word comprehension; evidence from nine fluent aphasic patients. *Aphasiology*, *3*(3), 189–207.

Fridriksson, J. (2010). Preservation and Modulation of Specific Left Hemisphere Regions is Vital for Treated Recovery from Anomia in Stroke. *Journal of Neuroscience*, *30*(35), 11558–11564.

Fridriksson, J., Bonilha, L., Baker, J. M., Moser, D., & Rorden, C. (2010). Activity in Preserved Left Hemisphere Regions Predicts Anomia Severity in Aphasia. *Cerebral Cortex*, *20*(5), 1013–1019.

Friederici, A. D., von Cramon, D. Y., & Kotz, S. A. (2007). Role of the Corpus Callosum in Speech Comprehension: Interfacing Syntax and Prosody. *Neuron*, *53*(1), 135–145.

Gajardo-Vidal, A., Lorca-Puls, D. L., Crinion, J. T., White, J., Seghier, M. L., Leff, A. P., … Price, C. J. (2018). How distributed processing produces false negatives in voxel-based lesion-deficit analyses. *Neuropsychologia*, *115*(March), 124–133.

Galliers, J., Wilson, S., Marshall, J., Talbot, R., Devane, N., Booth, T., … Greenwood, H. (2017). Experiencing EVA Park, a Multi-User Virtual World for People with Aphasia. *ACM Transactions on Accessible Computing*, *10*(4), 1–24.

Geva, S., Baron, J.-C., Jones, S. P., Price, C. J., & Warburton, E. A. (2012). *A comparison of {VLSM} and {VBM} in a cohort of patients with post-stroke aphasia*. *1*(1), 37–47.

Golestani, N., Paus, T., & Zatorre, R. J. (2002). Anatomical Correlates of Learning Novel Speech Sounds. *Neuron*, *35*, 997–1010. Retrieved from https://ac.els-cdn.com/S0896627302008620/1-s2.0-S0896627302008620-main.pdf?_tid=48fe51bf-77ac-483a-ba1c-41a3593153ab&acdnat=1549008449_9d3738a0904541e0ed011bc2446bebcb

Golestani, N., Price, C. J., & Scott, S. K. (2011). Born with an Ear for Dialects? Structural Plasticity in the Expert Phonetician Brain. *Journal of Neuroscience*, *31*(11), 4213–4220.

Good, C. D., Johnsrude, I. S., Ashburner, J., Henson, R., Friston, K. J., & Frackowiak, R. (2001). *A {Voxel-Based} Morphometric Study of Ageing in 465 Normal Adult Human Brains*. *14*(1), 21–36.

Goodglass, H., & Kaplan, E. (1972). *Assessment of aphasia and related disorders* (2nd ed). Beckenham : Philadelphia: Philadelphia: Lea & Febiger.

Grahn, J. A., Parkinson, J. A., & Owen, A. M. (2009). The role of the basal ganglia in learning and memory: Neuropsychological studies. *Behav Brain Res*, *199*(1), 53–60.

Grayson, E., Hilton, R., & Franklin, S. (1997). Early intervention in a case of jargon aphasia: efficacy of language comprehension therapy. *European Journal of Disorders of*

Communication : The Journal of the College of Speech and Language Therapists, London, *32*(1992), 257–276.

Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends in Neurosciences*, *25*(7), 348–353. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12079762

Grönholm, E. O., Roll, M. C., Horne, M. A., Sundgren, P. C., & Lindgren, A. G. (2016). Predominance of caudate nucleus lesions in acute ischaemic stroke patients with impairment in language and speech. *European Journal of Neurology*, *23*(1), 148–153.

Hamzei, F., Glauche, V., Schwarzwald, R., & May, A. (2012). Dynamic gray matter changes within cortex and striatum after short motor skill training are associated with their increased functional interaction. *NeuroImage*, *59*(4), 3364–3372.

Hervais-Adelman, A., Moser-Mercer, B., Murray, M. M., & Golestani, N. (2017a). Cortical thickness increases after simultaneous interpretation training. *Neuropsychologia*, *98*(February 2016), 212–219.

Hervais-Adelman, A., Moser-Mercer, B., Murray, M. M., & Golestani, N. (2017b). Cortical thickness increases after simultaneous interpretation training. *Neuropsychologia*, *98*(February 2016), 212–219.

Hessler, D., Jonkers, R., & Bastiaanse, R. (2010). *The influence of phonetic dimensions on aphasic speech perception*. *24*(12), 980–996.

Hickok, G. (2009). The functional neuroanatomy of language. *Physics of Life Reviews*, *6*(3), 121–143.

Hickok, G., Okada, K., Barr, W., Pa, J., Rogalsky, C., Donnelly, K., … Grant, A. (2008). Bilateral capacity for speech sound processing in auditory comprehension: Evidence from Wada procedures. *Brain and Language*, *107*(3), 179–184.

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

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*(1–2), 67–99.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402.

Hilari, K., Wiggins, R., Roy, P., Byng, S., & Smith, S. (2003). Predictors of health-related quality of life (HRQL) in people with chronic aphasia. *Aphasiology*, *17*(4), 365–381.

Hope, T. M. H., Leff, A. P., Prejawa, S., Bruce, R., Haigh, Z., Lim, L., … Price, C. J. (2017). Right hemisphere structural adaptation and changing language skills years after left hemisphere

stroke. *Brain*, *140*(6), 1718–1728.

Horowitz, A. L. (1995). *MRI physics for radiologists*. Springer: NY.

Howard, D., & Hatfield, F. M. (1987). *Aphasia Therapy - Historical and Contemporary Issues*. Lawrence Erlbaum Associates Ltd.: Hove, UK.

Howland, J., & Pierce, R. (2004). Influence of semantic relatedness and array size on single-word reading comprehension in aphasia. *Aphasiology*, *18*(11), 1005–1013.

Hugenschmidt, C. E., Peiffer, A. M., Kraft, R. A., Casanova, R., Deibler, A. R., Burdette, J. H., … Laurienti, P. J. (2008). Relating imaging indices of white matter integrity and volume in healthy older adults. *Cerebral Cortex*, *18*(2), 433–442.

Ijsselsteijn, W., Nap, H., de Kort, Y., & Poels, K. (2007). Digital Game Design for Elderly Users. *FuturePlay*.

Ilg, R., Wohlschlager, A. M., Gaser, C., Liebau, Y., Dauner, R., Woller, A., … Muhlau, M. (2008). Gray Matter Increase Induced by Practice Correlates with Task-Specific Activation: A Combined Functional and Morphometric Magnetic Resonance Imaging Study. *Journal of Neuroscience*, *28*(16), 4210–4215.

Jääskeläinen, I. P., Ahveninen, J., Belliveau, J. W., Raij, T., & Sams, M. (2007). Short-term plasticity in auditory cognition. *Trends in Neurosciences*, *30*(12), 653–661.

Jäncke, L., Hänggi, J., & Steinmetz, H. (2004). Morphological brain differences between adult stutterers and non-stutterers. *BMC Neurology*, *4*, 1–8.

Jette, D. U., Warren, R. L., & Wirtalla, C. (2005). The relation between therapy intensity and outcomes of rehabilitation in skilled nursing facilities. *Archives of Physical Medicine and Rehabilitation*, *86*(3), 373–379.

Kaas, J. H., & Hackett, T. A. (1999). “What” and “where” processing in auditory cortex. *Nature Neuroscience*, *2*(12), 1045–1047.

Kavanagh, M., Stanley, K., & Tyrrell, P. (2017). *Sentinel Stroke National Audit Programme (SSNAP) Rising to the Challenge: Stroke Care Received Between April 2016 to March 2017*. London: Healthcare Quality Improvement Partnership.

Kelley, W. M., Miezin, F. M., McDermott, K. B., Buckner, R. L., Raichle, M. E., Cohen, N. J., … Petersen, S. E. (1998). Hemispheric specialization in human dorsal frontal cortex and medial temporal lobe for verbal and nonverbal memory encoding. *Neuron*, *20*(5), 927–936.

Kiran, S., Sandberg, C., & Abbott, K. (2009). Treatment for lexical retrieval using abstract and concrete words in persons with aphasia: Effect of complexity. *Aphasiology*, *23*(7–8), 835–853.

Kiran, S., & Thompson, C. K. (2019). Neuroplasticity of Language Networks in Aphasia: Advances,

Updates, and Future Challenges. *Frontiers in Neurology*, *10*(April), 295.

Kleim, J. a, & Jones, T. a. (2008). Principles of experience-dependent neural plasticity: implications for rehabilitation after brain damage. *Journal of Speech, Language, and Hearing Research : JSLHR*, *51*(1), S225–S239.

Knecht, S., Drager, B., Deppe, M., Bobe, L., Lohman, H., Floel, A., … Henningsen, H. (2000). Handedness and hemispheric language dominance in healthy humans. *Brain*, *123*(12), 2512–2518.

Knollman-Porter, K., Dietz, A., & Dahlem, K. (2018). Intensive Auditory Comprehension Treatment for Severe Aphasia: A Feasibility Study. *American Journal of Speech-Language Pathology*, *27*(3), 936–949.

Krueger, R. A., & Casey, M. A. (2015). *Focus Groups - A Practical Guide for Applied Research* (5th ed.). Sage Publications Ltd. London.

Kurland, J., Wilkins, A., & Stokes, P. (2014). iPractice: Piloting the Effectiveness of a Tablet-Based Home Practice Program in Aphasia Treatment. *Seminars in Speech and Language*, *35*(01), 051–064.

Lambon Ralph, M. A., Pobric, G., & Jefferies, E. (2009). Conceptual knowledge is underpinned by the temporal pole bilaterally: Convergent evidence from rTMS. *Cerebral Cortex*, *19*(4), 832–838.

Lavoie, M., Macoir, J., & Bier, N. (2017). *Effectiveness of technologies in the treatment of post-stroke anomia: A systematic review*. *65*, 43–53.

Le Bihan, D. (2006). Looking into the functional architecture of the brain with diffusion MRI. *International Congress Series*, *1290*(June), 1–24.

Leech, R., Gygi, B., Aydelott, J., & Dick, F. (2009). Informational factors in identifying environmental sounds in natural auditory scenes. *The Journal of the Acoustical Society of America*, *126*(6), 3147–3155.

Leff, A. P., Crinion, J., Scott, S., Turkheimer, F., Howard, D., & Wise, R. (2002). A physiological change in the homotopic cortex following left posterior temporal lobe infarction. *Annals of Neurology*, *51*(5), 553–558.

Leff, A. P., Schofield, T. M., Crinion, J. T., Seghier, M. L., Grogan, A., Green, D. W., & Price, C. J. (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*(12), 3401–3410.

Li, J. P., Pan, P. L., Huang, R., & Shang, H. F. (2012). A meta-analysis of voxel-based morphometry studies of white matter volume alterations in Alzheimer's disease. *Neuroscience and*

Biobehavioral Reviews, *36*(2), 757–763.

Liakakis, G., Nickel, J., & Seitz, R. J. (2011). Diversity of the inferior frontal gyrus-A meta-analysis of neuroimaging studies. *Behavioural Brain Research*, *225*(1), 341–347.

Lohse, K. R., Lang, C. E., & Boyd, L. A. (2014). Is more better? Using metadata to explore dose-response relationships in stroke rehabilitation. *Stroke*, *45*(7), 2053–2058.

Lorca-Puls, D. L., Gajardo-Vidal, A., White, J., Seghier, M. L., Leff, A. P., Green, D. W., … Price, C. J. (2018). The impact of sample size on the reproducibility of voxel-based lesion-deficit mappings. *Neuropsychologia*, *115*(March), 101–111.

Lövdén, M., Wenger, E., Mårtensson, J., Lindenberger, U., & Bäckman, L. (2013). Structural brain plasticity in adult learning and development. *Neuroscience and Biobehavioral Reviews*, *37*(9), 2296–2310.

Lukic, S., Barbieri, E., Wang, X., Caplan, D., Kiran, S., Rapp, B., … Thompson, C. K. (2017). *Right Hemisphere Grey Matter Volume and Language Functions in Stroke Aphasia*. *2017*, 1–14.

Lumsden, J., Edwards, E. A., Lawrence, N. S., Coyle, D., & Munafò, M. R. (2016). Gamification of Cognitive Assessment and Cognitive Training: A Systematic Review of Applications and Efficacy. *JMIR Serious Games*, *4*(2), e11.

Luria, A. R., & Hutton, J. T. (1977). A modern assessment of the basic forms of aphasia. *Brain and Language*, *4*(2), 129–151.

Maldjian, J. A., Laurienti, P. J., Kraft, R. A., & Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *NeuroImage*, *19*(3), 1233–1239.

Mallet, K., Shamloul, R., Pugliese, M., Power, E., Corbett, D., Hatcher, S., … Dowlatshahi, D. (2019). RecoverNow: A patient perspective on the delivery of mobile tablet-based stroke rehabilitation in the acute care setting. *International Journal of Stroke*, *14*(2), 174–179.

Mandelli, M. L., Vilaplana, E., Brown, J. A., Hubbard, H. I., Binney, R. J., Attygalle, S., … Gorno-Tempini, M. L. (2016). Healthy brain connectivity predicts atrophy progression in non-fluent variant of primary progressive aphasia. *Brain*, *139*(10), 2778–2791.

Maneta, a, Marshall, J., & Lindsay, J. (2001). Direct and indirect therapy for word sound deafness. *International Journal of Language & Communication Disorders / Royal College of Speech & Language Therapists*, *36*(1), 91–106.

Manly, T., Davison, B., Heutink, J., Galloway, M., & Robertson, I. (2000). *Not enough time or not enough attention?: Speed, error and self-maintained control int he Sustained Attention to Response Test {(SART)}*. *3*, 167–177.

Marcotte, K., Daniel, A.-R., Damien, B., de Préaumont, M., Généreux, S., Hubert, M., & Ansaldo,

A. (2012). *Therapy-induced neuroplasticity in chronic aphasia*. *50*(8), 1776–1786.

Marcotte, K., Laird, L., Bitan, T., Meltzer, J. A., Graham, S. J., Leonard, C., & Rochon, E. (2018). Therapy-Induced Neuroplasticity in Chronic Aphasia After Phonological Component Analysis: A Matter of Intensity. *Frontiers in Neurology*, *9*, 225.

Marquié, J., Jourdan-Boddaert, L., & Huet, N. (2002). Do older adults underestimate their actual computer knowledge? *Behav Information Technology*, *21*(4), 273–280.

Marshall, J., Booth, T., Devane, N., Galliers, J., Greenwood, H., Hilari, K., … Woolf, C. (2016). *Evaluating the Benefits of Aphasia Intervention Delivered in Virtual Reality: Results of a {Quasi-Randomised} Study. 11*(8), e0160381.

Marshall, J. C., & Newcombe, F. (1966). *Syntactic and semantic errors in paralexia*. *4*(2), 169–176.

Marshall, J. C., & Newcombe, F. (1973). Patterns of paralexia: A psycholinguistic approach. *Journal of Psycholinguistic Research*, *2*(3), 175–199.

Marshall, J., Roper, A., Galliers, J., Wilson, S., Cocks, N., Muscroft, S., & Pring, T. (2013). Computer delivery of gesture therapy for people with severe aphasia. *Aphasiology*, *27*(9), 1128–1146.

Marston, H. R. (2013). Design Recommendations for Digital Game Design within an Ageing Society. *Educational Gerontology*, *39*(2), 103–118.

Mårtensson, J., Eriksson, J., Bodammer, N. C., Lindgren, M., Johansson, M., Nyberg, L., & Lövdén, M. (2012). Growth of language-related brain areas after foreign language learning. *NeuroImage*, *63*(1), 240–244.

Martin, A. M., Banks, C., Dietz, A. R., Ball, A., Allendorfer, J. B., Szaflarski, J. P., … Vannest, J. (2017). Neuroimaging Correlates of Post-Stroke Aphasia Rehabilitation in a Pilot Randomized Trial of Constraint-Induced Aphasia Therapy. *Medical Science Monitor*, *23*, 3489–3507.

Mcclelland, J. L., Fiez, J. A., & Mccandliss, B. D. (2002). Teaching the /r/–/l/ discrimination to Japanese adults: behavioral and neural aspects. *Physiology and Behavior*, *77*, 657–662. Retrieved from file:///Users/devans/Documents/Papers2/Articles/2003/Unknown/2003 R8705.pdf%5Cnpapers2://publication/uuid/D9D9D273-E580-4543-BB39-F6DA81E6B21F

McClenahan, R., Johnston, M., & Densham, Y. (1992). Factors influencing accuracy of estimation of comprehension problems in patients following cerebrovascular accident, by doctors, nurses and relatives. *International Journal of Language & Communication Disorders*, *27*(3), 209–219.

McKinnon, E. T., Fridriksson, J., Glenn, G. R., Jensen, J. H., Helpern, J. A., Basilakos, A., … Bonilha,

L. (2017). Structural plasticity of the ventral stream and aphasia recovery. *Annals of Neurology*, *82*(1), 147–151.

Mechelli, A., Crinion, J. T., Noppeney, U., O'Doherty, J., Ashburner, J., Frackowiak, R. S., & Price, C. J. (2004). Structural plasticity in the bilingual brain. *Nature*, *431*(7010), 757–757.

Mechelli, A., Price, C. J., Friston, K. J., & Ashburner, J. (2005). Voxel-Based Morphometry of the Human Brain: Methods and Applications. *Current Medical Imaging Reviews*, *1*(1).

Meinzer, M., Flaisch, T., Breitenstein, C., Wienbruch, C., Elbert, T., & Rockstroh, B. (2008). Functional re-recruitment of dysfunctional brain areas predicts language recovery in chronic aphasia. *NeuroImage*, *39*(4), 2038–2046.

Meinzer, M., Mohammadi, S., Kugel, H., Schiffbauer, H., Flöel, A., Albers, J., … Deppe, M. (2010). Integrity of the hippocampus and surrounding white matter is correlated with language training success in aphasia. *NeuroImage*, *53*(1), 283–290.

Meltzer, J. A., Wagage, S., Ryder, J., Solomon, B., & Braun, A. R. (2013). Adaptive significance of right hemisphere activation in aphasic language comprehension. *Neuropsychologia*, *51*(7), 1248–1259.

Menke, R., Meinzer, M., Kugel, H., Deppe, M., Baumgärtner, A., Schiffbauer, H., … Breitenstein, C. (2009). Imaging short- and long-term training success in chronic aphasia. *BMC Neuroscience*, *10*, 118.

Mestres-Missé, A., Càmara, E., Rodríguez-Fornells, A., Rotte, M., & Münte, T. (2008). Functional neuroanatomy of meaning acquisition from context. *Journal of Cognitive Neuroscience*, *20*(12), 2153–2166. Retrieved from http://www.mitpressjournals.org/doi/abs/10.1162/jocn.2008.20150

Miceli, G. (1980). Some aspects of phonological impairment in aphasia. *Brain and Language*, *11*(1), 159–169.

Miceli, Gabriele, Silveri, M. C., Nocentini, U., & Caramazza, A. (1988). Patterns of dissociation in comprehension and production of nouns and verbs. *Aphasiology*, *2*(3–4), 351–358.

Mills, K. L., & Tamnes, C. K. (2014). Methods and considerations for longitudinal structural brain imaging analysis across development. *Developmental Cognitive Neuroscience*, *9*, 172–190.

Minkova, L., Gregory, S., Scahill, R. I., Abdulkadir, A., Kaller, C. P., Peter, J., … Klöppel, S. (2018). Cross-sectional and longitudinal voxel-based grey matter asymmetries in Huntington's disease. *NeuroImage: Clinical*, *17*(June 2017), 312–324.

Mion, M., Patterson, K., Acosta-Cabronero, J., Pengas, G., Izquierdo-Garcia, D., Hong, Y. T., … Nestor, P. J. (2010). What the left and right anterior fusiform gyri tell us about semantic memory. *Brain*, *133*(11), 3256–3268.

Mohr, B. (2017). Neuroplasticity and Functional Recovery after Intensive Language Therapy in Chronic Post Stroke Aphasia: Which Factors Are Relevant? *Frontiers in Human Neuroscience*, *11*, 332.

Moineau, S., Dronkers, N. F., & Bates, E. (2005). Exploring the processing continuum of single-word comprehension in aphasia. *Journal of Speech, Language, and Hearing Research : JSLHR*, *48*(August), 884–896.

Morris, J., & Franklin, S. (2012). Investigating the effect of a semantic therapy on comprehension in aphasia. *Aphasiology*, *26*(12), 1461–1480.

Morris, J., Franklin, S., Ellis, A., Turner, J. E., & Bailey, P. J. (1996). *Remediating a speech perception deficit in an aphasic patient*. *10*(2), 137–158.

Mueller, J. L., Rueschemeyer, S. A., Ono, K., Sugiura, M., Sadato, N., & Nakamura, A. (2014). Neural networks involved in learning lexical-semantic and syntactic information in a second language. *Frontiers in Psychology*, *5*(OCT), 1–14.

Mummery, C. J., Ashburner, J., Scott, S. K., & Wise, R. J. S. (1999). Functional neuroimaging of speech perception in six normal and two aphasic subjects. *The Journal of the Acoustical Society of America*, *106*(1), 449–457.

Munro, P., & Siyambalapitiya, S. (2017). Improved word comprehension in Global aphasia using a modified semantic feature analysis treatment. *Clinical Linguistics & Phonetics*, *31*(2), 119–136.

Musso, M., Weiller, C., Kiebel, S., Müller, S. P., Bülau, P., & Rijntjes, M. (1999). Training-induced brain plasticity in aphasia. *Brain*, *122*(9), 1781–1790.

Naeser, M. A., Baker, E. H., Palumbo, C. L., Nicholas, M., Alexander, M. P., Samaraweera, R., … Weissman, T. (1998). Lesion site patterns in severe, nonverbal aphasia to predict outcome with a computer-assisted treatment program. *Archives of Neurology*, *55*(11), 1438–1448.

Nap, H. H., Kort, Y. A. W. De, & IJsselsteijn, W. A. (2009). Senior gamers: Preferences, motivations and needs. *Gerontechnology*, *8*(4).

Narain, C., Scott, S. K., Wise, R. J. S., Rosen, S., Leff, A., Iversen, S. D., & Matthews, P. M. (2003). Defining a Left-lateralized Response Specific to Intelligible Speech Using fMRI. *Cerebral Cortex*, *13*(12), 1362–1368.

Nardo, D., Holland, R., Leff, A. P., Price, C. J., & Crinion, J. T. (2017). Less is more: neural mechanisms underlying anomia treatment in chronic aphasic patients. *Brain*, *140*(11), 3039–3054.

Nickels, L. (2002). Theoretical and methodological issues in the cognitive neuropsychology of spoken word production. *Aphasiology*, *16*(1–2), 3–19.

Northcott, S., & Hilari, K. (2011). Why do people lose their friends after a stroke? *International Journal of Language & Communication Disorders / Royal College of Speech & Language Therapists*, *46*(5), 524–534.

O'Donnell, L. J., & Westin, C.-F. (2011). An Introduction to Diffusion Tensor Image Analysis. *Neurosurgery Clinics of North America*, *22*(2), 185–196.

Ohl, F. W., & Scheich, H. (2005). Learning-induced plasticity in animal and human auditory cortex. *Current Opinion in Neurobiology*, *15*(4), 470–477.

Oishi, K., Zilles, K., Amunts, K., Faria, A., Jiang, H., Li, X., … Mori, S. (2008). Human brain white matter atlas: Identification and assignment of common anatomical structures in superficial white matter. *NeuroImage*, *43*(3), 447–457.

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.-H., Saberi, K., … Hickok, G. (2010). Hierarchical Organization of Human Auditory Cortex: Evidence from Acoustic Invariance in the Response to Intelligible Speech. *Cerebral Cortex*, *20*(10), 2486–2495.

Palmer, R., Enderby, P., Cooper, C., Latimer, N., Julious, S., Paterson, G., … Hughes, H. (2012). Computer therapy compared with usual care for people with long-standing aphasia poststroke: A pilot randomized controlled trial (step by step). *Stroke*, *43*, 1904–1911.

Pani, E., Zheng, X., Wang, J., Norton, A., & Schlaug, G. (2016). Right hemisphere structures predict poststroke speech fluency. *Neurology*, *86*(17), 1574–1581.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*(12), 976–987.

Pedersen, P. M., Vinter, K., & Olsen, T. S. (2004). Aphasia after Stroke: Type, Severity and Prognosis. *Cerebrovascular Diseases*, *17*(1), 35–43.

Pike, C., Kritzinger, A., & Pillay, B. (2017). Social participation in working-age adults with aphasia: an updated systematic review. *Topics in Stroke Rehabilitation*, *24*(8), 627–639.

Pillay, S. B., Binder, J. R., Humphries, C., Gross, W. L., & Book, D. S. (2017). Lesion localization of speech comprehension deficits in chronic aphasia. *Neurology*, *88*(10), 970–975.

Plowman, E., Hentz, B., & Ellis, C. (2012). Post-stroke aphasia prognosis: A review of patient-related and stroke-related factors. *Journal of Evaluation in Clinical Practice*, *18*(3), 689–694.

Poeppel, D., & Hickok, G. (2004). Towards a new functional anatomy of language. *Cognition*, *92*(1–2), 1–12.

Pohl, A., Meyer, C. M., Deckers, A., Eckmann, T., Barthel, J., Huber, W., … Abel, S. (2017). Training-related changes of brain activation for speech production in healthy speakers–a longitudinal fMRI study to mimic aphasia therapy. *Aphasiology*, *31*(5), 495–518.

Pratchett, R. (2005). *Gamers in the UK: Digital play, digital lifestyles*. Retrieved from https://crystaltips.typepad.com/wonderland/files/bbc_uk_games_research_2005.pdf

Price, C. J. (1998). The functional anatomy of word comprehension and production. *Trends in Cognitive Sciences*, *2*(8), 281–288.

Price, C. J., Seghier, M. L., & Leff, A. P. (2010). Predicting language outcome and recovery after stroke: the {PLORAS} system. *Nat Rev Neurol*, *6*(4), 202–210.

Prins, R. S., Schoonen, R., & Vermeulen, J. (1989). *Efficacy of two different types of speech therapy for aphasic stroke patients*. *10*(1), 85–123.

Pulvermüller, F., & Berthier, M. L. (2008). Aphasia therapy on a neuroscience basis. *Aphasiology*, *22*(March 2015), 563–599.

Pulvermüller, F., Neininger, B., Elbert, T., Mohr, B., Rockstroh, B., Koebbel, P., & Taub, E. (2001). Constraint induced therapy of Aphasia After Stroke. *Stroke*, *32*, 2–7.

Rankin, E., Newton, C., Parker, A., & Bruce, C. (2014). Hearing loss and auditory processing ability in people with aphasia. *Aphasiology*, *28*(February 2015), 576–595.

Rauschecker, J. P. (2011). An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hearing Research*, *271*(1–2), 16–25.

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*(6), 718–724.

Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of ‘“what”’ and ‘“where”’ in auditory cortex. *PNAS*, *97*(22).

Raymer, A. M., & Kohen, F. (2006). Word-retrieval treatment in aphasia: Effects of sentence context. *Journal of Rehabilitation Research and Development*, *43*(3), 367–378.

Raymer, A. M., Kohen, F. P., & Saffell, D. (2006). Computerised training for impairments of word comprehension and retrieval in aphasia. *Aphasiology*, *20*(2–4), 257–268.

Recanzone, G. H., & Sutter, M. L. (2011). The biological basis of audition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(4), 408–418.

Rice, G. E., Caswell, H., Moore, P., Hoffman, P., & Lambon Ralph, M. A. (2018). The roles of left versus right anterior temporal lobes in semantic memory: A neuropsychological comparison of postsurgical temporal lobe epilepsy patients. *Cerebral Cortex*, *28*(4), 1487–1501.

Richardson, M. P., Strange, B. A., Duncan, J. S., & Dolan, R. J. (2003). Preserved verbal memory function in left medial temporal pathology involves reorganisation of function to right medial temporal lobe. *NeuroImage*, *20*(SUPPL. 1), 112–119.

Richter, M., Miltner, W. H. R., & Straube, T. (2008). *Association between therapy outcome and*

right-hemispheric activation in chronic aphasia. *131*(5), 1391–1401.

Robey, R. R. (1998). A Meta-Analysis of Clinical Outcomes in the Treatment of Aphasia. *Journal of Speech Language and Hearing Research*, *41*, 172–187.

Robles, G. S., Gatignol, P., Capelle, L., Mitchell, M.-C., & Duffau, H. (2005). The role of dominant striatum in language: a study using intraoperative electrical stimulations. *J Neurology Neurosurg Psychiatry*, *76*(7), 940–946.

Robson, H., Davies, S., Ralph, M. A., & Sage, K. (2012). *Facilitating and disrupting speech perception in word deafness*. *26*(2), 177–198.

Robson, H., Grube, M., Lambon Ralph, M. a., Griffiths, T. D., & Sage, K. (2013). Fundamental deficits of auditory perception in wernicke's aphasia. *Cortex*, *49*(7), 1802–1822.

Robson, H., Keidel, J. L., Lambon Ralph, M. A., & Sage, K. (2012). Revealing and quantifying the impaired phonological analysis underpinning impaired comprehension in Wernicke's aphasia. *Neuropsychologia*, *50*(2), 276–288.

Robson, H., Pilkington, E., Evans, L., DeLuca, V., & Keidel, J. L. (2017). Phonological and semantic processing during comprehension in Wernicke's aphasia: An N400 and Phonological Mapping Negativity Study. *Neuropsychologia*, *100*, 144–154.

Robson, H., Zahn, R., Keidel, J. L., Binney, R. J., Sage, K., & Ralph, M. A. (2014). *The anterior temporal lobes support residual comprehension in Wernicke's aphasia*. *137*(3), 931–943.

Roches, C. A., & Kiran, S. (2017). *Technology-Based Rehabilitation to Improve Communication after Acquired Brain Injury*. *11*, 382.

Rogalsky, C., Pitz, E., Hillis, A. E., & Hickok, G. (2008). Auditory word comprehension impairment in acute stroke: Relative contribution of phonemic versus semantic factors. *Brain and Language*, *107*(2), 167–169.

Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., & Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nature Neuroscience*, *2*(12), 1131–1136.

Rorden, C., & Brett, M. (2000). Stereotaxic display of brain lesions. *Behavioural Neurology*, *12*(4), 191–200.

Ross, K., & Wertz, R. (2003). Quality of life with and without aphasia. *Aphasiology*, *17*(4), 355–364.

San Chen, E., Han, M., Qi, Z., Gabrieli, J. D. E., & Garel, K. (2014). White-matter structure in the right hemisphere predicts Mandarin Chinese learning success. *Journal of Neurolinguistics*, *33*, 14–28.

Santos-Santos, M. A., Mandelli, M. L., Binney, R. J., Ogar, J., Wilson, S. M., Henry, M. L., … Gorno-

Tempini, M. L. (2016). Features of Patients With Nonfluent/Agrammatic Primary Progressive Aphasia With Underlying Progressive Supranuclear Palsy Pathology or Corticobasal Degeneration. *JAMA Neurology*, *73*(6), 733.

Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.-S., … Weiller, C. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(46), 18035–18040.

Saur, D., Lange, R., Baumgaertner, A., Schraknepper, V., Willmes, K., Rijntjes, M., & Weiller, C. (2006). *Dynamics of language reorganization after stroke*. *129*(6), 1371–1384.

Saur, D., Schelter, B., Schnell, S., Kratochvil, D., Küpper, H., Kellmeyer, P., … Weiller, C. (2010). Combining functional and anatomical connectivity reveals brain networks for auditory language comprehension. *NeuroImage*, *49*(4), 3187–3197.

Saygin, A. P., Dick, F., Wilson, S. W., Dronkers, N. F., & Bates, E. (2003). Neural resources for processing language and environmental sounds: Evidence from aphasia. *Brain*, *126*(4), 928–945.

Schlaug, G., Marchina, S., & Norton, A. (2009). Evidence for plasticity in white-matter tracts of patients with chronic broca's aphasia undergoing intense intonation-based speech therapy. *Annals of the New York Academy of Sciences*, *1169*, 385–394.

Schlegel, A. A., Rudelson, J. J., & Tse, P. U. (2012). White matter structure changes as adults learn a second language. *Journal of Cognitive Neuroscience*, *24*(8), 1664–1670.

Schmidt-Wilcke, T., Rosengarth, K., Luerding, R., Bogdahn, U., & Greenlee, M. W. (2010). Distinct patterns of functional and structural neuroplasticity associated with learning Morse code. *NeuroImage*, *51*(3), 1234–1241.

Schneider, E. J., Lannin, N. A., Ada, L., & Schmidt, J. (2016). Increasing the amount of usual rehabilitation improves activity after stroke: a systematic review. *Journal of Physiotherapy*, *62*(4), 182–187.

Scholz, J., Klein, M. C., Behrens, T. E. J., & Johansen-Berg, H. (2009). Training induces changes in white-matter architecture. *Nature Neuroscience*, *12*(11), 1370–1371.

Scott, S. K. (2012). The neurobiology of speech perception and production-Can functional imaging tell us anything we did not already know? *Journal of Communication Disorders*, *45*(6), 419–425.

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123 Pt 12*, 2400–2406.

Seghier, M. L., Ramlackhansingh, A., Crinion, J., Leff, A. P., & Price, C. J. (2008). Lesion identification using unified segmentation-normalisation models and fuzzy clustering.

NeuroImage, *41*(4), 1253–1266.

Seghier, M. L., Ramsden, S., Lim, L., Leff, A. P., & Price, C. J. (2014). Gradual lesion expansion and brain shrinkage years after stroke. *Stroke*, *45*(3), 877–879.

Sexton, C. E., Walhovd, K. B., Storsve, A. B., Tamnes, C. K., Westlye, L. T., Johansen-Berg, H., & Fjell, A. M. (2014). Accelerated Changes in White Matter Microstructure during Aging: A Longitudinal Diffusion Tensor Imaging Study. *Journal of Neuroscience*, *34*(46), 15425–15436.

Shtyrov, Y. (2012). Neural bases of rapid word learning. *Neuroscientist*, *18*(4), 312–319.

Stahl, B., Mohr, B., Büscher, V., Dreyer, F. R., Lucchese, G., & Pulvermüller, F. (2018). Efficacy of intensive aphasia therapy in patients with chronic stroke: a randomised controlled trial. *Journal of Neurology, Neurosurgery & Psychiatry*, *89*(6), 586–592.

State of the Nation, Stroke Statistics. (2017). Stroke Association.

Stebbins, G. T., Nyenhuis, D. L., Wang, C., Cox, J. L., Freels, S., Bangen, K., … Gorelick, P. B. (2008). Gray Matter Atrophy in Patients With Ischemic Stroke With Cognitive Impairment. *Stroke*, *39*(3), 785–793.

Stoodley, C. J., & Schmahmann, J. D. (2009). Functional topography in the human cerebellum: A meta-analysis of neuroimaging studies. *NeuroImage*, *44*(2), 489–501.

Storey, P. (2006). Introduction to Magnetic Resonance Imaging and Spectroscopy. In P. V. Prasad (Ed.), *Magnetic Resonance Imaging. Methods in Molecular Medicine™* (Vol. 124, pp. 3–57).

Swinburn, K., Howard, D., & Porter, G. (2004). *CAT : comprehensive aphasia test*. Hove; New York: Psychology Press.

Szabo, G., & Dittelman, J. (2014). Using mobile technology with individuals with aphasia: Native iPad features and everyday apps. *Seminars in Speech and Language*, *35*(1), 5–16.

Tardif, C. L., Gauthier, C. J., Steele, C. J., Bazin, P. L., Schäfer, A., Schaefer, A., … Villringer, A. (2016). Advanced MRI techniques to improve our understanding of experience-induced neuroplasticity. *NeuroImage*, *131*, 55–72.

Tessier, C., Agnes, W.-C., Michelot, N., & Pascale, P.-D. (2007). Rehabilitation of word deafness due to auditory analysis disorder. *Brain Injury*, *21*(11), 1165–1174.

The MathWorks, I. (2017). *MATLAB and Statistics Toolbox Release 2017a.* Natick, Massachusetts, United States.

Thomas, A. G., Marrett, S., Saad, Z. S., Ruff, D. A., Martin, A., & Bandettini, P. A. (2009). Functional but not structural changes associated with learning: An exploration of longitudinal Voxel-Based Morphometry (VBM). *NeuroImage*, *48*(1), 117–125.

Thomas, C., & Baker, C. I. (2013). Teaching an adult brain new tricks: A critical review of evidence

for training-dependent structural plasticity in humans. *NeuroImage*, *73*, 225–236.

Thompson, C. K., Choy, J. W. J., Holland, A., & Cole, R. (2010). Sentactics®: Computer-automated treatment of underlying forms. *Aphasiology*, *24*(10), 1242–1266.

Thompson, H. E., Robson, H., Lambon Ralph, M. A., & Jefferies, E. (2015). Varieties of semantic 'access' deficit in Wernicke's aphasia and semantic aphasia. *Brain*, *138*(12), 3776–3792.

Tian, B., Reser, D., Durham, A., Kustov, A., Josef, P., Kenyon, C., … Fridovich, I. (2001). Functional Specialization in Rhesus Monkey Auditory Cortex Published by : American Association for the Advancement of Science Functional Specialization in Rhesus Monkey Auditory Cortex. *Science*, *292*, 290–293.

Turkeltaub, P. E., Messing, S., Norise, C., & Hamilton, R. H. (2011). Are networks for residual language function and recovery consistent across aphasic patients? *Neurology*, *76*(20), 1726–1734. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed10&NEWS=N&AN=2011288592

Turken, A. U., & Dronkers, N. F. (2011). The Neural Architecture of the Language Comprehension Network: Converging Evidence from Lesion and Connectivity Analyses. *Frontiers in System Neuroscience*, *5*, 1.

Vallila-Rohter, S., & Kiran, S. (2013). Non-linguistic learning and aphasia: Evidence from a paired associate and feedback-based task. *Neuropsychologia*, *51*(1), 79–90.

Van Hees, S., McMahon, K., Angwin, A., de Zubicaray, G., & Copland, D. A. (2014). Neural activity associated with semantic versus phonological anomia treatments in aphasia. *Brain and Language*, *129*(1), 47–57.

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology (2006)*, *67*(February 2015), 1176–1190.

van Oers, C. A. M. M., Vink, M., van Zandvoort, M. J. E., van der Worp, H. B., de Haan, E. H. F., Kappelle, L. J., … Dijkhuizen, R. M. (2010). Contribution of the left and right inferior frontal gyrus in recovery from aphasia. A functional MRI study in stroke patients with preserved hemodynamic responsiveness. *NeuroImage*, *49*(1), 885–893.

Varley, R. (2011). Rethinking aphasia therapy: A neuroscience perspective. *International Journal of Speech-Language Pathology*, *13*(1), 11–20.

Varley, R., Cowell, P. E., Dyson, L., Inglis, L., Roper, A., & Whiteside, S. P. (2016). Self-Administered Computer Therapy for Apraxia of Speech: Two-Period Randomized Control Trial With Crossover. *Stroke*, *47*(3), 822–828.

Veekmans, K., Nopp, P., D'Haese, P., & Moeltner, A. (2004). Speech comprehension over the telephone in adult cochlear implant subjects. *Cochlear Implants International*, *5*(SUPPL. 1), 94–96.

Vigneau, M., Beaucousin, V., Hervé, P. Y., Duffau, H., Crivello, F., Houdé, O., … Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage*, *30*(4), 1414–1432.

Vigneau, M., Beaucousin, V., Hervé, P. Y., Jobard, G., Petit, L., Crivello, F., … Tzourio-Mazoyer, N. (2011). What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? Insights from a meta-analysis. *NeuroImage*, *54*(1), 577–593.

Visch-brink, E. G., Stronks, D. L., & Denes, G. (2005). *De semantische associatie test*. Harcourt Test Publishers.

Wagner, A. D., Prull, M. W., Glover, G. H., Gabrieli, J. D. E., Desmond, J. E., & Poldrack, R. A. (2002). Functional Specialization for Semantic and Phonological Processing in the Left Inferior Prefrontal Cortex. *NeuroImage*, *10*(1), 15–35.

Walenski, M., Europa, E., Caplan, D., & Thompson, C. K. (2019). Neural networks for sentence comprehension and production: An ALE-based meta-analysis of neuroimaging studies. *Human Brain Mapping*, (December 2018), 1–30.

Wan, C. Y., Zheng, X., Marchina, S., Norton, A., & Schlaug, G. (2014). Intensive therapy induces contralateral white matter changes in chronic stroke patients with Broca's aphasia. *Brain and Language*, *136*, 1–7.

Warren, J. E., Wise, R. J. S., & Warren, J. D. (2005). Sounds do-able: Auditory-motor transformations and the posterior temporal plane. *Trends in Neurosciences*, *28*(12), 636–643.

Watila, M. M., & Balarabe, S. A. (2015). Factors predicting post-stroke aphasia recovery. *Journal of the Neurological Sciences*, *352*(1–2), 12–18.

Webster, J., Whitworth, A., & Morris, J. (2015). Is it time to stop "fishing"? A review of generalisation following aphasia intervention. *Aphasiology*, *29*(11), 1240–1264.

Wernicke, C. (1908). The Symptom-Complex of Aphasia. In *Diseases of the nervous system* (pp. 265–323). D. Appleton and Company: Newy York and London.

Whitworth, A., Webster, J., & Howard, D. (2014). *A Cognitive Neuropsychological Approach to Assessment and Intervention in Aphasia* (2nd ed.). Psychology Press.

Wilson, S. M. (2017). Lesion-symptom mapping in the study of spoken language understanding. *Language, Cognition and Neuroscience*, *32*(7), 891–899.

Wilson, S., Roper, A., Marshall, J., Galliers, J., Devane, N., Booth, T., & Woolf, C. (2015). Codesign

for people with aphasia through tangible design languages. *CoDesign*, *11*(1), 21–34.

Wise, R. J. S., Scott, S. K., Blank, S. C., Mummery, C. J., Murphy, K., & Warburton, E. A. (2001). Separate neural subsystems within `Wernicke's area'. *Brain*, *124*(1), 83–95.

Woodhead, Z., Kerry, S. J., Aguilar, O. M., Ong, Y.-H. H., Hogan, J. S., Pappa, K., … Crinion, J. T. (2018). Randomized trial of iReadMore word reading training and brain stimulation in central alexia. *Brain*, *141*(7), 2127–2141.

Woodhead, Z. V., Crinion, J., Teki, S., Penny, W., Price, C. J., & Leff, A. P. (2017). Auditory training changes temporal lobe connectivity in 'Wernicke's aphasia': a randomised trial. *Journal of Neurology, Neurosurgery & Psychiatry*, *88*(7), 586–594.

Woolf, C., Panton, A., Rosen, S., Best, W., & Marshall, J. (2014). Therapy for auditory processing impairment in aphasia: An evaluation of two approaches. *Aphasiology*, *28*(12), 1481–1505.

Xing, S., Lacey, E. H., Skipper-Kallal, L. M., Jiang, X., Harris-Love, M. L., Zeng, J., & Turkeltaub, P. E. (2016). Right hemisphere grey matter structure and language outcomes in chronic left hemisphere stroke. *Brain*, *139*(1), 227–241.

Xing, S., Lacey, E. H., Skipper-Kallal, L. M., Zeng, J., & Turkeltaub, P. E. (2017). White matter correlates of auditory comprehension outcomes in chronic post-stroke aphasia. *Frontiers in Neurology*, *8*(FEB), 1–9.

Young, S. (1984). *Nuclear Magnetic Resonance Imaging - Basic Principles*. Raven Press: NY.

Zatorre, R. J., Fields, R. D., & Johansen-Berg, H. (2012). Plasticity in gray and white: Neuroimaging changes in brain structure during learning. *Nature Neuroscience*, *15*(4), 528–536.

Zheng, C., Lynch, L., & Taylor, N. (2015). Effect of computer therapy in aphasia: a systematic review. *Aphasiology*, 1–34.

8 Appendices

Appendix 1 – Focus Group Two format

Focus group 2

Format

- Two small groups ideally separated into different rooms due to significant impact of background noise during FG1.
- Each participant to have a tablet with headphones. Participants will each be assigned, and sign for, their own tablet. It is important that participants only use their designated tablet, and that these do not get mixed up between participants.

Brief Schedule

1. Introduction →Connect
2. Very brief demo to group on how to access the app on the tablet →Sonia/Victoria
3. 5 minutes reminding participants of earlier FG1 prototype version →Sonia/Victoria
4. 5-10 minutes playing new FG2 prototype
5. Main focus group session with questions →Connect
6. Demo and independent play and feedback on telephone level, and background noise level →Connect
7. Very brief demo of chapters part of app →Sonia/Victoria
8. Discussion about taking the tablet home →Victoria/Sonia

Each of the following questions to be supplemented by:

- Screenshots on A4 for:
    - Each question
    - Each Visual Analogue Scale
    - Comparisons of old/new elements
    - →Sonia/Victoria to email these to Connect
- Visual Analogue Scales, with:
    - A screenshot for the part of the app the rating scale refers to
    - No numbers or markers on the line
    - → Connect
- Any further visual aids to support total communication → Connect

General notes from FG1

- Encourage independent play as much as possible with using the tablets and app (e.g. which buttons to press). We want to see how participants use the game if they don't have anyone to help them. This will help us decide if we need to make changes e.g. adding stickers to buttons on the tablet.
- Encourage equal feedback between all members as much as possible, so that members with expressive difficulties are equally represented in our data. We has lots of great feedback in the last group, but a bit less from some members compared to others, so want to make feedback as equal as possible.
- Supported and non-verbal feedback is good too, as our videos can capture this. We want to make sure all members, regardless of their particular communication difficulty, are represented in our data.
- VAS scale:
    - These worked really well in FG1, so we want to continue using these in FG2.
    - Can we have a plain VAS scale without numbers or markers. If people prefer saying a number out of 10 that is fine, we can record that down instead.

Questions

1. **We have added a "fire" button to the pinball game. Do you prefer having a "fire" button, or pressing anywhere on the screen?**

Screenshots: old version with no button, new version with button
Multiple choice: old / new

2. **Do you prefer overlapping pictures, or separate pictures?**

Screenshots: old version without overlapping, new version with overlapping
Multiple choice: old / new

3. **Do you understand the pinball game?**
   (*This is a key question, so we can spend lots of time discussing this)

Probe: do members understand the link between:
winning coins → "spending" the coins in the pinball game → winning a jigsaw piece → completing a puzzle.
Screenshots: one for each stage of the game above
VAS scale: "Do you understand the pinball game and jigsaw pieces?" (No-------Yes)

4. **Do you like playing the pinball game?**

Probe: what could make it better/more fun?
Screenshots: pinball game
VAS scale: “Do you enjoy playing the pinball game?” (No-------Yes)

5. **Do you like how the new jigsaw puzzles <u>look</u>?**

Screenshots: old jigsaw puzzle screens, new jigsaw puzzle screens
VAS scale: “Do you like how the jigsaw puzzle looks?” (No--------Yes)

6. **Would you enjoy collecting jigsaw pieces, to finish the puzzle?**

Screenshots: winning jigsaw piece in pinball game, jigsaw puzzle screen
VAS scale: “Do you like collecting jigsaw pieces to finish the puzzle?” (No--------Yes)

7. **Was the audio okay in the game? How was the volume? How was the quality?**

Screenshots: challenge screen
VAS scale: “Are the voices in the game ok?” (No------Yes)

8. **How comfortable are the headphones?**

(No resources needed)

9. **Introduce the telephone level, 5-10 minutes playing time**

Questions:
Did you notice a change in the voice? (multiple choice: yes/no)
What did you think?

Screenshots: telephone level

10. **Introduce background noise level, 5-10 minutes playing time**

Questions:
Did you notice a change in what you heard? (multiple choice: yes/no)
What did you think?

Screenshots: background noise level

11. **Introduce chapter element of game (no questions on this needed)**

12. **Introduce taking tablet home**

Aim: to play for at least 10 minutes every day
5 minutes for participants to ask questions about taking the tablet home.

Appendix 2 – Self-report measure: subtests and questions

Expressive

1. This set of questions ask about **you talking to** different people

''In the past month how easy was it for you….?''

a. To talk to your wife/husband/partner/children?
b. To talk to friends?
c. To talk to a stranger?
d. To talk on the telephone?

Receptive

2. The next set of questions ask about when different people are **talking to you.**

''In the past month how easy was it for you….?''

a. To understand your wife/husband/partner/children when they were talking to you?
b. To understand friends when they were talking to you?
c. To understand a stranger when they were talking to you?
d. To understand in a conversation in a group?
e. To understand in a conversation in a noisy place?
f. To understand somebody on the telephone?

Daily tasks

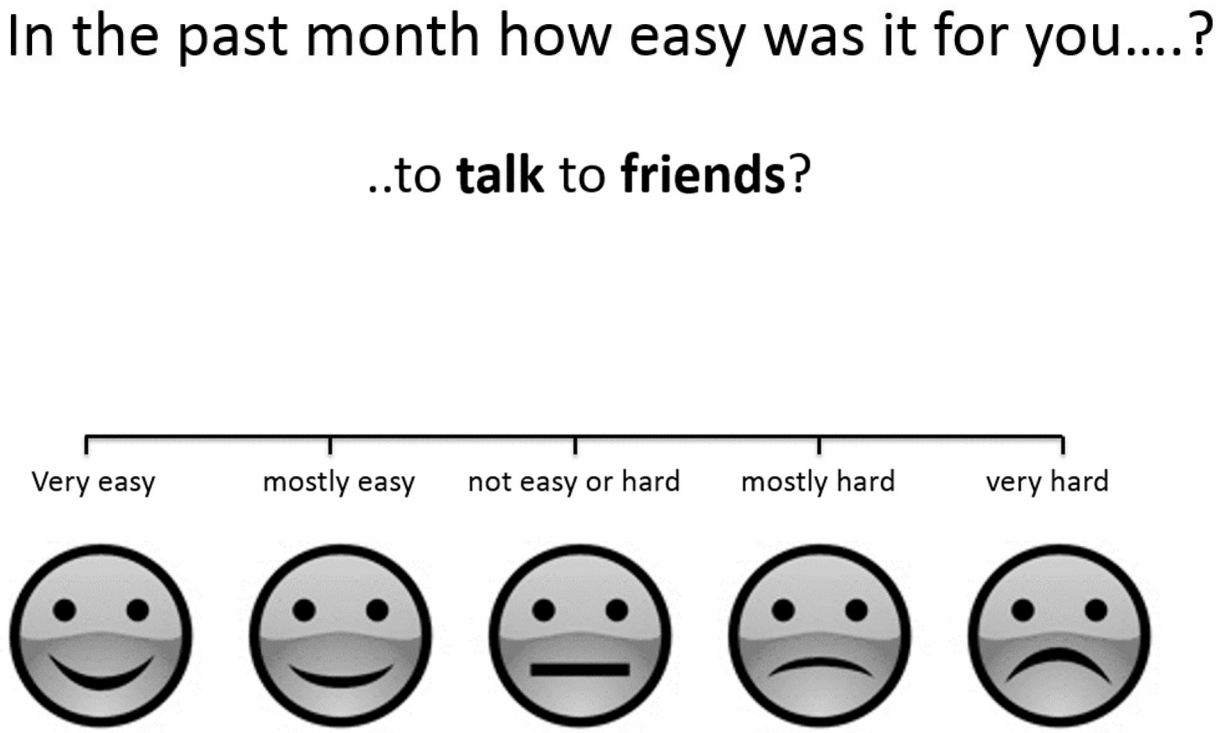
3. The next set of questions ask about **doing** different tasks

''In the past month how easy was it for you….?''

a. To write a note (i.e. a shopping list, a card)?
b. To read a newspaper article?
c. To communicate with other people overall?
d. To do fun things/hobbies?
e. To find things around the home?
f. Wash and dress yourself?
g. Prepare a meal or snack?
h. Move around without bumping into things.

Example question-response format

Patients were required to point to a face.



Scoring

Points were awarded for each question as follows:

very easy=1,

mostly easy=2,

not easy/not hard=6,

mostly hard=4,

very hard=5.

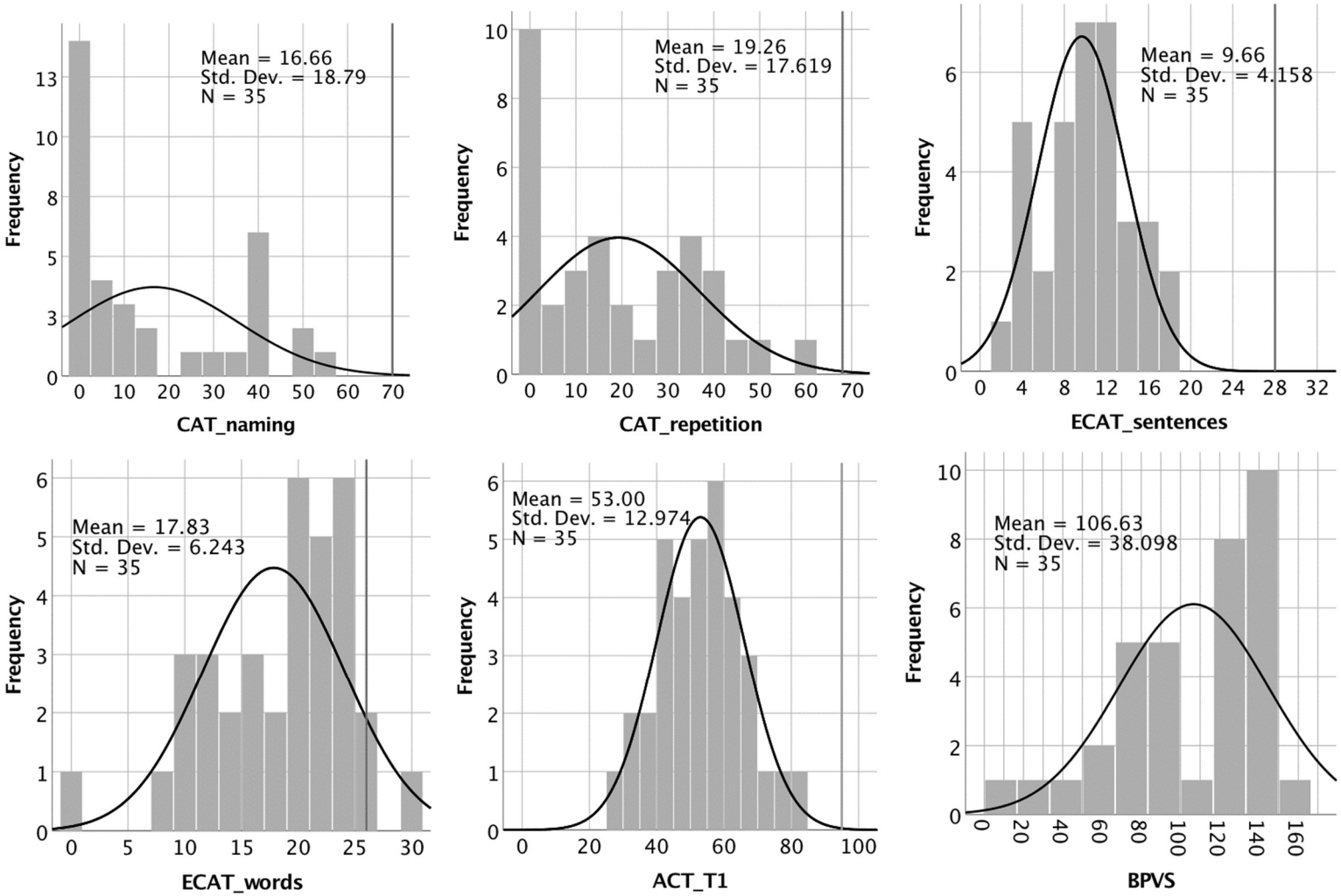
Total points possible for each subsection were:

Expressive language=20,

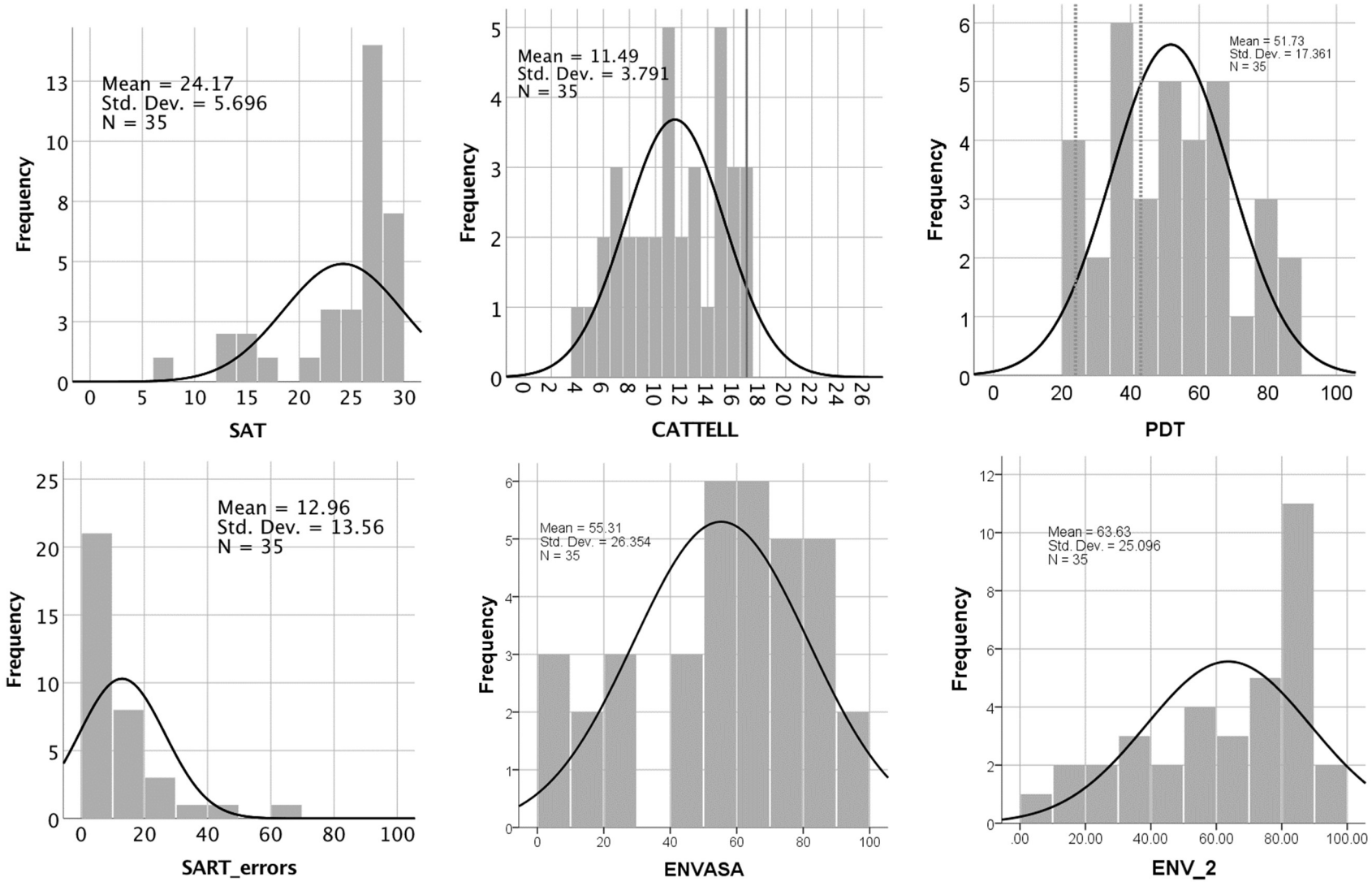
Receptive language=30,

Control tasks=40.

Appendix 3 – Histograms of subtest performance at T1 (N=35)



Red line indicates aphasia cut off scores for CAT-naming, CAT-repetition, CAT-sentences and CAT-words; and performance by age matched controls for ACT (N=22), CATTELL (N=27). PDT: scores within dotted red lines indicate chance level performance (p>.05)



Red line indicates aphasia cut off scores for CAT-naming, CAT-repetition, CAT-sentences and CAT-words; and performance by age matched controls for ACT (N=22), CATTELL (N=27). PDT: scores within dotted red lines indicate chance level performance (p>.05)

Appendix 4 – Performance on subtests of Comprehensive Aphasia Test at T1 (N=35)