

Neural and Cognitive Mechanisms Affecting Perceptual Adaptation to Distorted Speech

Dan Kennedy-Higgins

Speech, Hearing and Phonetic Sciences

Thesis submitted for the degree of Doctor of Philosophy

University College London

June 2018

Primary Supervisor: Dr Patti Adank

Secondary Supervisor: Dr Joseph T. Devlin

Declaration

I, Dan Kennedy-Higgins 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.

Dan Kennedy-Higgins

Abstract

The majority of everyday communication occurs in the presence of distortions, such as background noise, yet the human ability to understand speech in adverse listening conditions is remarkably robust. Past research has investigated perceptual adaptation to different speech conditions, however, our knowledge of the individual differences and the associated cognitive and neural mechanisms affecting perceptual adaptation is still limited. The work described in this thesis therefore aimed to advance our understanding of this research area, with specific focus first on determining the extent to which adaptation to one distortion generalises to another, second, determining the underlying cognitive mechanisms of this adaptation process and finally determining what role, if any, the left ventral premotor cortex plays in adaptation.

This thesis presents results from eight experiments, two behavioural and six using Transcranial Magnetic Stimulation (TMS) as the primary research tool. Results from experiments 1 and 2 (behavioural) show that measures of verbal intelligence, specifically vocabulary knowledge, working memory and general cognitive functioning underpin the perceptual learning process, providing support for statistical learning to occur and assist adaptation to distorted speech. Additionally, the results suggest participants possess a general skill that enables generalisation of learning from one adverse listening condition to another. Experiments 3 to 8 used TMS to modulate perception of speech in noise in a bilateral superior temporal region. However, no effect of using this protocol was found when applied to the left ventral premotor cortex whilst participants adapted to time-compressed speech.

The results of the experiments described in this thesis are considered in the context of our current understanding of the cognitive and neural mechanisms associated with perceptual adaptation to distorted speech. It is believed that the results will contribute significantly to existing knowledge due to use of novel research methodologies e.g., use of multiple distortions, multiple speakers and TMS.

Impact Statement

Experiments 1 and 2 presented within the second chapter of this thesis are among first to investigate the cognitive mechanisms supporting perceptual adaptation to distorted speech using multiple speech distortions and multiple speakers in a within participant design. Outside of second language learning research, previous work has not systematically explored the relationship between speaker and distortion. As a result, it was impossible to determine whether adaptation processes were truly independent of the speaker's idiosyncratic vocal characteristics. The results presented in chapter two suggest that these two factors (distortion and vocal characteristics) interact. This has potentially important ramifications firstly for the field as a whole, as the majority of studies use a single speaker during adaptation. It would be impossible to establish whether the presence or absence of an adaptation effect was due to the characteristics of the distortion, the speaker or both. Second, and more importantly beyond the laboratory, noise-vocoded speech is believed to simulate the experience of using a cochlear implant. If adaptation is dependent on vocal characteristics as well as distortion, then this would suggest that there are certain individuals for whom cochlear implant users will find it explicitly harder to perceive and adapt to. This finding therefore has the potential to impact on the design and fitting of auditory aids such as cochlear implants. In future, the design of such aids will need to consider how the idiosyncrasies of different speakers are perceived by the listener, and, where possible, which characteristics need to be manipulated in order to maximise speaker intelligibility.

In addition, the experiments presented within the third and fourth chapters of this thesis use Transcranial Magnetic Stimulation (TMS) to investigate the neural mechanisms affecting perceptual adaptation to distorted speech. Within academia, the results of the experiments presented in chapter three are particularly relevant as they present the effects of four different TMS protocols on performance of the same speech perception task. The results of these experiments may help to inform other researchers of the potential issues faced and the best practices to adopt when designing TMS experiments with the aim of investigating the neurobiology of speech perception. Finally, TMS affords the ability to investigate brain-behaviour relationships in a transient but causal way. This allows observation of immediate effects of changing normal brain function, and thus provides an insight into the types of behavioural and

cognitive impairment that would be expected following clinical damage, e.g., due to strokes or lesions. By having a better idea of the types of deficits faced following damage to specific regions TMS allows for the development of more efficacious treatment and strategies to be developed and implemented. It is believed that the results of this thesis can contribute to our pre-existing knowledge and has the potential to inform choices in clinical, as well as laboratory settings.

Table of Contents

Chapter One.....	20
Introduction	20
Research Aims.....	22
Literature Review	22
Anatomy and Physiology of the Auditory Pathway	22
Anatomy and Physiology of Speech Perception: the Dorsal and Ventral Functional Streams	27
Neural Responses to Distorted Speech.....	31
Neural Mechanisms of Auditory Perceptual Adaptation	35
Transcranial Magnetic Stimulation	36
Outline of the Thesis	38
Chapter Two	39
Adaptation to Distorted Speech: Cognitive Mechanisms and the Effect of Listening Distortion and Speaker Variability	39
Introduction	39
Time-Compressed Speech.....	40
Noise-Vocoded Speech	48
Speech in Noise	53
Transfer of Learning from one Distortion to Another.....	56
Transfer of Learning from one Speaker to Another	57
Individual Differences in Perceptual Adaptation	59
Summary of research aims	65
Experiment One.....	65
Methods.....	65
Participants	65
Procedure.....	65

Audiological Assessments	66
Cognitive Assessments	66
Speech Adaptation Task.....	68
Data Analysis	70
Results	70
Accuracy	70
Response Times	71
Correlation in Performance Between Conditions	74
Analysis of Speaker Acoustic Characteristics.....	77
Relationship Between Cognitive Assessments and Performance in Different Speech Conditions.....	78
Interim Summary	83
Experiment Two.....	83
Methods.....	83
Participants.....	83
Procedure.....	84
Results	84
Accuracy	84
Response Times	85
Benefit of Multiple Speakers per Condition: A Comparison of Results from Experiment One and Two	87
Discussion	88
Transfer of Learning from one Condition to Another and the Associated Cognitive Mechanisms.....	89
Perceptual Adaptation to the Different Speech Distortions	91
Transfer of Learning from one Speaker to Another.....	94
Conclusions	97
Chapter Three	99

Investigating the Efficacy of TMS as a Research tool to Disrupt Auditory Sentential Processing (in Noise)	99
Introduction	99
Experiment Three	110
Methods	110
Participants	110
Procedure.....	110
Stimuli	112
Transcranial Magnetic Stimulation	112
Selection of TMS sites	114
MRI scanning	115
Data Analysis	115
Results	117
Interim discussion	120
Experiment Four.....	121
Methods.....	121
Participants	121
Transcranial Magnetic Stimulation	121
Procedure.....	122
Results	123
Interim discussion	126
Experiment Five	127
Methods.....	127
Participants	127
Transcranial Magnetic Stimulation	128
Procedure.....	128
Results	129

Interim Discussion	131
Experiment Six.....	132
Methods.....	132
Participants.....	132
Test of Etymotic ER1 earphones.....	132
Transcranial Magnetic Stimulation.....	133
Procedure.....	134
Results	134
Interim discussion	136
Experiment Seven	138
Methods.....	138
Participants.....	138
Procedure.....	138
MNI-152 structural brain scan	140
Results	141
Experiments Six and Seven.....	143
Results	143
General Discussion.....	143
Conclusions	150
Chapter Four	151
Investigating the Role of the Ventral Premotor Cortex in Adaptation to Time-Compressed Speech.	151
Introduction.....	151
Experiment Eight	159
Methods.....	159
Participants.....	159
Procedure.....	159

Stimuli	160
Transcranial Magnetic Stimulation	161
Site of Stimulation.....	162
Data Analysis	163
Results	164
Accuracy.....	164
Response Times.....	168
Discussion	172
Conclusions	175
Chapter Five	177
General Discussion.....	177
Summary of research aims	177
Review of the originality and impact of presented work and discussion of limitations and future research directions	178
Generalisation of learning across adverse listening conditions and the underlying cognitive mechanisms (related to the first and second research aims).....	178
Effect of exposure to multiple talkers during adaptation (related to the third research aim)	182
Investigating the most effective TMS protocol to non-invasively impair sentential speech perception in healthy human adults (related to the fourth research aim)	186
Role of the left ventral premotor cortex in adaptation to time-compressed speech (related to the fifth research aim)	187
Conclusions	189
Summary of the significant findings of this thesis.....	190
Bibliography	191
List of Appendices	208

List of Figures

Figure 1 - Accuracy and Response Times for Experiment One across Condition, Speaker and Sentence Number.....	73
Figure 2 - Scatterplots displaying the relationship between accuracy of performance in each of the adverse listening conditions. Solid line represents the line of best fit. Dotted lines represent 95 percent confidence intervals.....	75
Figure 3 - Scatterplots displaying the relationship between response times performance in each of the adverse listening conditions. Solid line represents the line of best fit. Dotted lines represent 95 percent confidence intervals.	76
Figure 4 - Accuracy and Response Times for Experiment Two across Condition, Speaker and Sentence Number.....	86
Figure 5 - Illustration of the order of events and associated timings for one cortical site in Experiment Three.	114
Figure 6 - Boxplots displaying SRT performance across the different TMS stimulation sites for Experiment Three.	118
Figure 7 - Boxplots displaying GDT performance across the different TMS stimulation sites for Experiment Three.	119
Figure 8 - Illustration of the order of events and associated timings for one cortical site in Experiment Four.	122
Figure 9 - Boxplots displaying SRT performance across the different TMS stimulation sites for Experiment Four.....	124
Figure 10 - Boxplots displaying GDT performance across the different TMS stimulation sites for Experiment Four.....	125
Figure 11 - Boxplots displaying SRT performance across the different TMS stimulation sites for Experiment Five.	129
Figure 12 - Boxplots displaying SRT performance across the different TMS stimulation sites for Experiment Six.	135
Figure 13 - Illustration of a single trial and associated timings for Experiment Seven.	139
Figure 14 - Boxplots displaying Visual Discrimination threshold performance across the different TMS stimulation sites for Experiment Seven.....	142
Figure 15 - Boxplots displaying z-scores across the different TMS stimulation sites for Experiments Six and Seven.	143

Figure 16 - A - Illustration of the order of events and associated timings for Experiment Eight. B - Illustration of a single trial and associated timings of TMS pulses for Experiment Eight. Vertical dashes represent each occurrence of a TMS pulse. 160

Figure 17 - A - Line graphs displaying Accuracy of performance for Experiment Eight across TMS group, Condition and Block Number. B - Line graphs displaying Accuracy of performance for Experiment Eight across Condition and Block Number. 166

Figure 18 - Line graphs displaying Accuracy of performance for Experiment Eight across TMS group, Condition, Block Number and Order of Condition. 167

Figure 19 - A - Line graphs displaying Response Times performance for Experiment Eight across TMS group, Condition and Block Number. B - Line graphs displaying Response Time performance for Experiment Eight across Condition and Block Number..... 169

Figure 20 - Line graphs displaying Response Times performance for Experiment Eight across TMS group, Condition, Block Number and Order of Condition..... 170

List of Tables

Table 1 - Summary of the previous research investigating the link between distorted speech and cognitive mechanisms.	62
Table 2 - Acoustic Characteristics of the different speakers.	69
Table 3 - Statistics from individual Spearman’s rank-order correlations on accuracy data.	75
Table 4 - Statistics from individual Spearman’s rank-order correlations on response data.	76
Table 5 - Bonferroni corrected Mann-Whitney U pairwise comparison tests of Median average f_0 values.	77
Table 6 - Bonferroni corrected Mann-Whitney U pairwise comparison tests of median syllable rate.	78
Table 7 - Descriptive statistics for the battery of audiological and cognitive assessments.	78
Table 8 - Rotated Structure Matrix for PCA with Varimax Rotation.	79
Table 9 - Summary of Multiple Regression Analyses for Accuracy Data across conditions.	81
Table 10 - Summary of Multiple Regression Analyses for Response Time data across conditions.	82
Table 11 - Showing means and standard deviations for accuracy data across each condition and experiment.	87
Table 12 - Showing means and standard deviations for response time data across each condition and experiment.	88
Table 13 - Summary of previous research studies investigating the effect of TMS to temporal cortical regions.	106
Table 14 - Average group co-ordinates across experiments.	115
Table 15 - Summary of group descriptive statistics for SRT task across TMS condition.	118
Table 16 - Summary of group descriptive statistics for GDT task across TMS condition.	119
Table 17 - Summary of group descriptive statistics for SRT task across TMS condition.	124

Table 18 - Summary of group descriptive statistics for GDT task across TMS condition.....	126
Table 19 - Summary of group descriptive statistics for SRT task across TMS condition.....	130
Table 20 - Summary of the pairwise comparison statistics for experiment six.....	135
Table 21 - Summary of group descriptive statistics for SRT task across TMS condition.....	142
Table 22 - Summary Statistics of the four different groups in Experiment Eight..	159
Table 23 - Means and Standard Deviations of the Accuracy and Response Time data for the two speech condition across blocks of 15 sentences.	171

List of Abbreviations

A1	Primary field of Primary Auditory Cortex
AF	Arcuate Fasciculus
ALE	Activation Likelihood Estimation
aMT	Active Motor Threshold
ANOVA	Analysis of Variance
aSTG	Anterior Superior Temporal Gyrus
BOLD	Blood-Oxygen Level Dependent
CV	Consonant-Vowel
dB	Decibels
EEG	Electroencephalography
EMG	Electromyography
f_0	Fundamental Frequency
F1	First Formant
F2	Second Formant
FDI	First Dorsal Interosseous
fMRI	Function Magnetic Resonance Imaging
GDT	Gap Detection Threshold
GUI	Graphical User Interface
H_0	Null Hypothesis
H_1	Alternative Hypothesis
HG	Heschl's Gyrus
HL	Hearing Level
Hz	Hertz
IC	Inferior Colliculus
IEEE	Institute of Electrical and Electronics Engineers
IFG	Inferior Frontal Gyrus
IFOF	Inferior Frontal-Occipital Fasciculus
JND	Just Noticeable Difference
LOC	Lateral Occipital Complex
LogMAR	Logarithm of the Minimum Angle of Resolution
IPMv	Left Ventral Premotor
MEG	Magnetoencephalography
MEPs	Motor Evoked Potentials
MGN	Medial Geniculate Nucleus
mm	Millimetres
MMNm	Magnetic Mismatch Negativity
MNI	Montreal Neurological Institute
MoCA	Montreal Cognitive Assessment
MRI	Magnetic Resonance Imaging
msecs	Milliseconds
MTG	Middle Temporal Gyrus
MTSP	Motor Theory of Speech Perception
PAC	Primary Auditory Cortex
PACT	Perception for Action Control Theory

PC	Personal Computer
PCA	Principal Component Analysis
PET	Positron Emission Tomography
PPA	Primary Progressive Aphasia
PSOLA	Pitch Synchronous Overlap and Add
pSTG	Posterior Superior Temporal Gyrus
PT	Planum Temporale
PTA	Pure Tone Audiometry
R	Rostral field of Primary Auditory Cortex
RGB	Red Green Blue
RHT	Reverse Hierarchy Theory
RTf	Rostral-Temporal field of Primary Auditory Cortex
RT	Response Times
rTMS	Repetitive Transcranial Magnetic Stimulation
SCOLP	Speech and Capacity of Language Processing
SD	Standard Deviation
SE	Standard Error
SI	Intermediate Sulcus
SNR	Signal-to-Noise Ratio
SPL	Sound Pressure Level
SRT	Speech Recognition Threshold
STG	Superior Temporal Gyrus
STS	Superior Temporal Sulcus
sPMv	Superior Ventral Premotor
syl/sec	Syllables per Second
TMS	Transcranial Magnetic Stimulation
TMT	Trail-Making Test
UCL	University College London
UF	Uncinate Fasciculus
PMv	Ventral Premotor

Acknowledgments

First, I must acknowledge Patti for everything over the past five years. I couldn't have asked for a better supervisor. For letting me stretch the rules, for always challenging me to do better, for letting me disappear for six months, for all the advice and for all the meetings where anything other than the most obvious subject was discussed, I am eternally grateful and enjoyed every second (well, nearly every second!). Thanks to Helen for being the best postdoc a PhD student could ever wish to have. I miss our long, rambling scientific discussions that ultimately solved nothing but were still somehow incredibly useful. To Gwijde, Shego, Julie, Adam and Max for so politely entertaining my many mumbling scientific rants, for treating me with such undeserved esteem and for making the lab such a happy place to be.

Thanks to Joe for being a phenomenal second supervisor, your words of advice and generosity of time and knowledge have always been greatly appreciated. We'll never know how much more we could have achieved if the time wasn't spent discussing football.... but how boring the last seven years would have been otherwise! Thanks to John, Tae, Magda and Sylvia for making me realise how much I love research and for setting a standard that I've tried to attain ever since (but have never quite achieved).

Merci Beaucoup to Pascale, Isabelle, Melody, Didier and Catherine for going so far beyond the necessary limits to make me welcome in your corner of the world, there are not enough words to express my gratitude.

Thanks to Axelle for helping me see the magniblimful world beyond the lab, from Montreal to Machu Picchu, Highgate to Hekla and beyond. Thanks to Hao, Faith, Gise, Jieun, Yue, Emma, Tim, Outi, Joaquín and Andrew for making Chandler a great place to work, for always asking how life is and for appearing to care what the answer was. Thanks to Stuart, Jyrki, Carol, Rachel and Merle for always looking out for me and for providing more than a fair share of encouragement and support along the way.

Thanks to all of the participants that gave up their brains to make this research possible and to Beth and Jane for taking the time to read this thesis, for trying to find its strengths and hopefully looking kindly upon its (many) weaknesses.

Thanks to Ken Bruce, Jeremy Vine, Steve Wright, Simon Mayo, Gordon Lundy, Loz Guest, Mark Church, Johnny Barren, Adam Buxton, Desert Island Discs, Just a Minute, The News Quiz, Radiolab and Counting Crows for providing the soundtrack to my PhD and for keeping me (relatively) sane during the countless hours spent coding, writing, analysing and pondering whether I'd misunderstood all along and p really was supposed to be greater than 0.05.

Last but most importantly, thanks to Ja for never allowing others to ask how the thesis is going; to Dave for the many, many happy evenings spent at the Oval (and some less happy days spent north of the river); to Nourane for the film, TV and book recommendations that provided a great escape from work. To Squaz (my Godmother!) and Michael without whom I may never have been drawn to psychology. I thank you for every influence, I grew up trying to imitate you both, I think I turned out alright.

Thanks to Dug, who told me these would be the results five years ago, if only I'd listened. To Georgie, you ran your marathon and went back for more, I think I'll stop at one. Finally, to Grace, my partner in the museum of non-significant results, you've been there for me every step of the way and have provided me with the perfect reason to not spend my life at work, I'll forgive you for not taking part in any of my experiments!

This thesis is dedicated to Ma and Pa, the two greatest parents in the world, who encourage me to follow my dreams with unwavering love and support. I love you both.

“Science, my lad, is made up of mistakes, but they are mistakes which it is useful to make, because they lead little by little to the truth.”

Jules Verne, *A Journey to the Centre of the Earth*

Chapter One

Introduction

The overarching aim of this thesis is to develop a more comprehensive understanding of the cognitive and neural mechanisms affecting perceptual adaptation to distorted speech. A speech distortion refers to any context in which the ease of perception has been reduced to the point where successful comprehension is not guaranteed. Mattys, Davis, Bradlow, and Scott (2012) define three main types of speech distortion. (1) Production related distortions occur due to a number of factors including, coarticulation, and differences between speakers in accent, age and gender (and thus vocal tract shape and size) or due to the effects of neurogenic disorders of production such as dysarthria or structural abnormalities of the articulators such as cleft palate. (2) Environmental related distortions most commonly occur as a result of energetic or informational masking from rival signals, for example, competing talkers in a café or background noise at a busy train station. (3) Perceiver related distortions can occur as a result of anatomical changes, such as, sensorineural hearing impairment or a stroke/lesion resulting in central neurological perception deficits such as aphasia. Additionally, individual differences in cognitive functioning are also associated with differences in successful perception of speech.

Despite such an array of distortions prevalent in everyday life, speech is still perceived with relative ease and minimal conscious effort on the part of the (healthy) perceiver. Even with exaggerated levels of alteration, due to simulated distortions such as noise-vocoded speech (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995); or artificially time-compressed speech (Dupoux & Green, 1997) listeners have repeatedly shown an ability to adapt to the challenges of the particular distortion after just a few minutes of exposure. Yet the processes that underlie this ability, especially the neural processes by which adaptation to the distorted speech occurs, are still poorly understood.

As stated above, the primary aim of this thesis is to add to our understanding of the cognitive and neural mechanisms involved in adapting to speech presented in different adverse listening conditions. In order to do so this thesis adopts several unique methodological approaches. The two experiments presented in chapter two

investigate the way in which a single set of ninety participants adapt to three different forms of speech distortion (time-compressed, noise-vocoded and speech in noise) using the recordings from four different adult male speakers. This is in contrast to previous research that has predominantly investigated perceptual adaptation using a single distortion and/or speaker. By investigating the rates of adaptation to multiple distortions in a single set of participants, this thesis aims to determine whether individuals possess a general ability to adapt to any kind of adverse listening condition or whether they use distortion dependent adaptation strategies. For example, are certain individuals particularly good at adapting to spectral manipulations but poor at adapting to temporal manipulations, or is the type of distortion irrelevant? By using recordings from multiple speakers, I investigated the extent to which adaptation is dependent on the spectral and temporal characteristics of the distortion, the vocal idiosyncrasies of the speaker, or a combination of the two. In addition, the experiments presented in chapter two aim to expose the underlying cognitive mechanisms associated with individual differences in adaptation to each of the three adverse listening conditions. Research linking individual differences in perceptual adaptation with different underlying cognitive mechanisms (e.g., vocabulary knowledge, working memory) is becoming increasingly popular and the results from experiment one and two of this thesis are expected to contribute significantly to this burgeoning field of research.

Furthermore, previous research investigating the neural mechanisms underlying perceptual adaptation to distorted speech has almost exclusively used functional imaging techniques e.g. functional magnetic resonance imaging (fMRI). Whilst the information gained from this research is undoubtedly of critical importance, the conclusions drawn are limited by the correlational nature of associating changes in blood oxygen levels in a given cortical region to ongoing task performance. Therefore, the final unique methodological approach of this thesis is to adopt Transcranial Magnetic Stimulation (TMS) as the primary research tool used in chapters three and four. In these chapters, TMS is used to investigate the role of the left and right superior temporal sulcus (STS) in the perception of speech in noise (chapter three) and the role of the left ventral premotor (PMv) cortex and superior temporal sulcus (STS) in adaptation to time-compressed speech (chapter four). It is believed that the use of TMS

will enable more definitive conclusions to be drawn related to the role of different cortical structures in adaptation to distorted speech stimuli.

Research Aims

1. To determine the extent to which learning of one type of speech distortion generalises to the learning of other types of speech distortion.
2. To expose the underlying cognitive mechanisms associated with individual differences in adaptation.
3. To determine the extent to which exposure to multiple speakers impacts overall adaptation.
4. To find the most effective TMS protocol to non-invasively impair accurate perception of speech in noise in healthy human adults.
5. To investigate the role of the left ventral premotor cortex in adaptation to time-compressed speech.

Literature Review

Before introducing the hypothesised cognitive and neural underpinnings of adaption to distorted speech, the following sections will, first, outline how sounds in general are processed by central and peripheral auditory structures. Second, introduce the neuroanatomy of speech perception in both clear and distorted listening conditions and third, outline the neuroanatomical structures involved in adaptation to distorted speech.

Anatomy and Physiology of the Auditory Pathway

The initial action of the auditory system is to transduce pressure waves in the air into mechanical movements of the ear before subsequent translation into electrical energy in the auditory nerve. All of these processes occur in the peripheral auditory system, which is made up of the outer, middle and inner ear. The outer ear consists of the pinna (the visible part of the ear) and the auditory canal. The role of the outer ear is to gather

sound from the fluctuating pressure waves around us and direct it towards the tympanic membrane and middle ear (Plack, 2013; Rosen & Howell, 2011).

The middle ear functions to reduce the mismatch in impedances between the air filled outer ear and the fluid filled inner ear. The pressure waves gathered by the outer ear cause the tympanic membrane to move in and out. The movement of this membrane results in the movement of the interconnected ossicular bones and in turn the oval window. Whilst the outer and middle ear fulfil very important functions, these functions are generally considered to 'enable' hearing whilst the actual process of 'hearing' is believed to take place in the inner ear (Plack, 2013). The mechanical energy transmitted through the middle ear is converted into pressure fluctuations of the cochlea fluid in the inner ear. The distance along the basilar membrane that the pressure fluctuations propagate depends on the wavelength/frequency of transduced waves. High frequencies propagate a short distance causing movement of the first parts of the cochlea, whilst low frequencies propagate the length of the basilar membrane towards the apex (Robles & Ruggero, 2001). Sat on top of the basilar membrane is the Organ of Corti which contains up to six rows of hair cells, five of which have outer hair cells which are hypothesised to affect the motion of the basilar membrane, whilst one row contains inner hair cells which are thought to be responsible for converting the vibration of the basilar membrane into electrical activity in the auditory nerve. When the basilar membrane and the tectorial membrane move up and down relative to one another this causes the stereocilia on the hair cells to sway from side to side. The stereocilia of each hair cell are connected by filaments known as tip links, when the stereocilia are bent towards the scala media the tip links are stretched. This stretching causes them to pull on tiny trap doors blocking channels in the membrane of the stereocilia. When these channels open up, positively charged potassium ions flow into the hair cell causing it to depolarise. This depolarisation triggers the release of glutamate neurotransmitters into the synaptic cleft between the hair cell and the neuron of the auditory nerve (Hudspeth, 2014). A larger movement of the basilar membrane results in more tip links being opened and more neurotransmitter being released into the synaptic cleft and thus greater electrical activity in the auditory nerve. As each inner hair cell is attached to a specific place on the basilar membrane, the activated auditory nerve carries information about the vibration of the basilar membrane at a single point in the cochlea. The tonotopic mapping of the basilar membrane is therefore

replicated on the auditory nerve, where very specific parts of the nerve respond to narrow frequency ranges, a concept referred to as tuning. It has been proposed that this tuning is the single most important function in the neural perception of speech sounds (Young, 2008). Without this frequency-specific tuning, it is likely that only the single most intense part of a sound would be perceived with all other spectral detail lost. For example, the second formant of a vowel would not be perceived as it would be masked by the more intense first formant. However, as a result of the frequency specific tuning, different parts of the auditory nerve respond to specific frequencies thus decomposing sounds by their strongest frequencies and maintaining the integrity of the complex spectral structure of the sound, thereby facilitating a more accurate neural representation.

Before the auditory information reaches the Primary Auditory Cortex (PAC) it is processed in a number of places in the brainstem. All axons of the auditory nerve end in the cochlear nuclei, where information related to the location of the sound source is initially gathered, as well as an initial improvement in the signal to noise ratio (SNR). From the cochlear nucleus, the majority of neurons project via the lateral lemniscus to the inferior colliculus (IC), which is the first major acoustic processing hub in the midbrain, whilst a subset of neurons from the anterior-ventral cochlear nucleus take a more indirect route to the IC via the superior olivary complex where the information from both ears comes together for the first time. Nerve fibres in the IC then synapse with the medial geniculate nucleus (MGN) in the thalamus. The MGN has three main subdivisions: ventral, dorsal and medial nuclei. The ventral subdivision maintains the same tonotopic organisation as all preceding structures and projects predominantly to the primary auditory region of the cerebral cortex, whilst the dorsal and medial structures lack a clear tonotopic organisation (Schnupp, Nelken, & King, 2011). The dorsal and medial MGN structures project to both primary and non-primary auditory cortex, with the majority of dorsal neurons projecting to secondary auditory regions (Helfert, Snead, & Altschuler, 1991).

At present our knowledge of the exact architecture of human PAC is incomplete however it is believed to closely resemble that of non-human primates where extensive anatomical research has been conducted. In monkeys, 13 sub-regions have been identified as belonging to the primary auditory cortex with these regions combining to make up three major primary regions: the core, belt and parabelt. The

core encompasses three sub-regions (Primary field A1; Rostral field R; Rostral-Temporal field RTf) each of which have dense individual and reciprocal connections with the medial geniculate nucleus in the thalamus. The individual or parallel nature of these connections is important as it ensures that ablation of one core region does not result in complete deactivation of the other two (Kaas & Hackett, 2000). In addition to the thalamic connections, each core region interconnects heavily with its neighbouring core region, adjacent belt regions and unlike primary visual and somatosensory regions, the core also has substantial interhemispheric connections with the cytoarchitecturally corresponding core/belt regions in the contralateral hemisphere. All of which suggests that the core regions both heavily influence each other, as well as the surrounding belt and contralateral core/belt regions.

Whilst the core regions do project transcallosally to their corresponding contralateral core and belt regions, the vast majority of within hemisphere projections are to the surrounding and adjacent belt regions. Very few projections exist from the core to the parabelt regions and no cortical regions beyond the parabelt have direct connections to the core. This profile of connections suggests a step-wise manner of processing with the core performing initial processing and the belt conducting secondary level of cortical processing subsequent to the core.

In the human brain, the primary auditory cortex (PAC) occupies most of the transverse temporal gyrus, known as Heschl's gyrus (HG), deep within the lateral sulcus (Sylvian Fissure) of each hemisphere. HG is bordered medially by the insular cortex, laterally by the superior temporal gyrus, anteriorly by the first transversal sulcus and posteriorly by Heschl's sulcus (S. Clarke & Morosan, 2012). In humans, this region displays anatomical individual differences and can consist of between one and three gyri per person, per hemisphere with the number of gyri per hemisphere not necessarily being equal. The anatomical variance is partially due to the path of the intermediate sulcus (SI). In some individuals, this sulcus does not intersect HG and results in a single smooth gyrus whilst in others the SI can either partially or fully divide HG resulting in two parallel gyri (Da Costa et al., 2011; Penhune, Zatorre, MacDonald, & Evans, 1996). The PAC, as with primary sensory and visual areas, can be identified cytoarchitecturally based on its well-developed inner-granular layer (layer 4). From an evolutionary point of view, Heschl's gyrus is a recent structure being present in chimpanzee brains but not in the macaque monkey (Hackett, Preuss,

& Kaas, 2001; Moerel, De Martino, & Formisano, 2014). Despite the recent evolutionary development of Heschl's gyrus and therefore the lack of its presence in closely related species, Heschl's gyrus is believed to be the site of the human homologue of the core area of PAC. In non-human primates, each of the three regions of the core (A1, R and RTf) show tonotopic gradients that are mirror symmetrical of each other and each respond well and with short latencies to pure tones. Using high field strength magnetic resonance imaging similar tonotopic functional organisation has been found to exist along Heschl's gyrus in humans. Using silent event-related fMRI, Formisano et al. (2003) found a region of HG progressing from a caudal to rostral location that showed a graded blood oxygen level dependent (BOLD) response to stimuli of different frequencies. More caudal regions exhibited greater preferences to high frequency tones with more rostral regions showing preference to low frequency tones. With an adjacent, and more rostral region showing the mirror opposite response, these two symmetrical gradient areas are believed to be the human homologue of core areas A1 and R found in the non-human primate anatomy. Humphries, Liebenthal, and Binder (2010) also found frequency-selective regions on the supratemporal plane of human subjects with one region extending posterior-medially from HG towards Planum Temporale, and the second region extending anterior-medially to an area close to the first transverse sulcus (Moerel et al., 2014) with the most posterior-medial or anterior-medial areas showing preferential responses to high frequency tones, these results complement those of Formisano et al. (2003). In addition, Humphries et al. (2010) found evidence of a third region in the posterior lateral region of the superior temporal gyrus (STG) which illustrated a smaller gradient and could represent the human equivalent of core region RTf (Da Costa et al., 2011; Morosan et al., 2001; Wessinger, Buonocore, Kussmaul, & Mangun, 1997). Formisano et al. (2003) also found evidence of clusters outside of the two mirror symmetrical areas within HG where the BOLD response to the pure tones was weaker and less specific, it is suggested that this could represent activation within the human homologue of the belt region of PAC.

In non-human primate research, the belt has been shown to comprise eight sub-regions that surround the core region of PAC. This is in accordance with research in human anatomy that has found evidence of several granular fields surrounding the core that have less dense cell packing and larger, more voluminous pyramidal cells in layer

three (S. Clarke & Morosan, 2012; Moerel et al., 2014). In monkeys, the belt is heavily interconnected with both core and parabelt regions but has fewer thalamic connections. The few connections the belt does have with the thalamus terminate in the dorsal MGN as opposed to the core which receives inputs mainly from ventral MGN. In addition to the neighbouring core and parabelt regions, the sub-areas of the belt also connect to more distant belt regions, i.e., not only to adjacent regions, with some evidence to suggest that the belt regions of PAC in monkeys may project to areas of the prefrontal cortex. Such connections would represent the first connections in the auditory processing chain outside of classic primary auditory regions. As suggested above, evidence from both human and non-human primate research suggests that a hierarchy of processing occurs within the PAC with core regions responding to pure tones and belt regions responding to more complex sounds e.g. bandpass noise bursts. Wessinger et al. (2001) used fMRI to specifically investigate the core-belt model of hierarchical auditory processing, the researchers found that both pure tones and band-pass noise activated core regions of bilateral PAC. However, activation related to pure tones did not extend beyond a well-defined core region, with areas surrounding this core region only activated by band-pass noise and not by pure tones, thus supporting both the notion of the human equivalent to the non-human primate core-belt dissociation and a hierarchy in the processing of complex sounds.

Finally, the parabelt is divided into two sub-regions, both of which are located on the lateral side of the lateral belt. In non-human primates, the parabelt receives the vast majority of its inputs from the adjacent belt areas with few connections from the auditory core. Importantly in the hierarchical processing of core to belt to parabelt, this region is interconnected with several regions of the temporal, parietal and frontal lobes of the monkey brain, representing the real starting point of more higher level acoustic analyses and sound recognition processes.

Anatomy and Physiology of Speech Perception: the Dorsal and Ventral Functional Streams

The processing of all sounds (speech or otherwise) is believed to be equal and bilateral up to the PAC with this region showing sensitivity to surface acoustics e.g. changes in frequency whereas the neural processes associated with higher level abstraction e.g.

sound identification and lexical access, occur beyond the primary auditory cortex. These higher-level regions need to bind lower level acoustic features (that develop rapidly over short time windows) together to form complex spectrotemporal forms. In the process of forming the complex forms, the higher-level regions need to overcome the acoustic variance of individual sounds and thus need to show an invariant response to natural stimulus variation. It is through this combination of low level acoustic features into higher more complex acoustic objects that we are able to overcome differences in pronunciation due to accent, age, gender or acoustic environment and instead perceive the complex spectrotemporal forms as categorical phonemes, words and/or utterances (DeWitt & Rauschecker, 2012; Scott & Johnsrude, 2003).

A key feature in higher level processing within the context of a hierarchical model would therefore be regions of the brain that respond in a consistent manner to speech stimuli despite the vast amount of acoustic variation that occurs. This is important as it would suggest that such regions are responding to the categorical linguistic information in the signal rather than the acoustic idiosyncrasies (Pelle, Johnsrude, & Davis, 2010). Evidence for such regions comes from Davis and Johnsrude (2003) who compared the BOLD response to three different speech conditions which were matched in levels of intelligibility but differed acoustically. The authors used noise-vocoded speech, speech in noise and speech segmented by bursts of noise with each form of distortion presented at three levels of intelligibility (i.e., low, medium, and high). It was found that activation within PAC did not correlate reliably with intelligibility, i.e., the PAC responded to all acoustic stimuli equally, highlighting its role in processing low level purely acoustic information. Differential levels of activation specific to the type of stimuli but irrespective of intelligibility (form-dependent regions) occurred within the belt and parabelt regions of auditory cortex supporting the notion that these regions are sensitive to differences in surface acoustic structure but are not involved in higher level and more abstract processes. Whereas the opposite pattern of activation, that is, activation specifically correlated with intelligibility but insensitive to acoustic variation was observed in bilateral anterior STG, left posterior STG and left inferior frontal gyrus. Each of these regions became more active as the level of intelligibility increased across all speech conditions, suggesting that these regions are involved in processing speech at more abstract, non-acoustic levels of representation. Subsequent evidence to support the role of bilateral

anterior and left dominant posterior regions of superior temporal cortex and inferior frontal gyrus in higher level abstract linguistic processing has been reported in numerous studies (Evans, Kyong, Rosen, Golestani, Warren, McGettigan, Mourao-Miranda, et al., 2014; Friederici, Kotz, Scott, & Obleser, 2010; Harris, Dubno, Keren, Ahlstrom, & Eckert, 2009; Narain et al., 2003; Okada et al., 2010; Rosen, Wise, Chadha, Conway, & Scott, 2011; Scott, Blank, Rosen, & Wise, 2000; Zekveld, Heslenfeld, Festen, & Schoonhoven, 2006). Whilst it is possible and likely that each of these regions is capable of performing the same abstract function (Price & Friston, 2002), evidence in support of two distinct functional streams in spoken language comprehension suggests that the different regions perform predominantly different functions (Hickok & Poeppel, 2004; Rauschecker & Scott, 2009).

Extending out from the primary auditory cortex there are two main functional streams of processing, referred to as the dorsal and ventral streams. The dorsal stream initially projects posteriorly and dorsally from the PAC towards the inferior parietal cortex before bending around towards posterior inferior frontal regions. The dominant task of this stream is to map sounds onto articulatory motor maps with the most prototypical task of this stream to enable vocal repetition of perceived sounds (Saur et al., 2008; Selnes, Knopman, Niccum, & Rubens, 1985; Warren, Wise, & Warren, 2005). Instrumental to the functioning of the dorsal stream is the Arcuate Fasciculus, a white matter tract that connects the temporal lobe to the posterior inferior frontal regions. The tract consists of three segments (in both hemispheres). The traditional pathway, referred to as the direct segment, arches around the lateral fissure and connects posterior temporal regions to the middle and inferior frontal and ventral premotor cortex. The posterior segment connects posterior superior and middle temporal regions to inferior parietal regions whilst the anterior segment connects inferior parietal regions to middle and inferior frontal regions and ventral premotor cortex. The anterior and posterior segments run parallel and lateral to the direct segment (Catani, Jones, & ffytche, 2005; Catani & Thiebaut de Schotten, 2012; Thiebaut de Schotten et al., 2011). Damage to this tract is associated with Conduction Aphasia, a language deficit which is characterised by an impaired ability to repeat utterances and paraphasic production, despite (relatively) spared perceptual abilities. It is suggested that this deficit is due to an impairment in the combining of phonological word forms with the associated sequential articulatory gestures

(Tomasino et al., 2015). Furthermore, direct stimulation of the AF in awake neurosurgical patients produces phonological paraphasias (Maldonado, Moritz-Gasser, & Duffau, 2011; Mandonnet, Nouet, Gatignol, Capelle, & Duffau, 2007) representative of the phonemic production errors found in patients (Ardila, 1992) and the microstructural integrity of the tract is linked with phonological awareness (Yeatman et al., 2011) and reading ability in children (Deutsch et al., 2005; Niogi & McCandliss, 2006) and pseudoword language learning in adults (López-Barroso et al., 2013) all of which supports the hypothesis that the dorsal stream is predominantly involved in phonological processing and the mapping of phonological forms onto their associated articulatory actions.

In contrast, the ventral processing stream predominantly underlies the mapping of sound to semantic meaning. That is, the ventral stream is hypothesised to incorporate all of the functional processes, beyond the PAC, required to extract meaning from the incoming auditory stimuli in order to comprehend what is being heard. Anatomically, this stream projects in a lateral and ventral direction away from the PAC to the middle and inferior temporal cortices extending to the temporal pole and/or inferior frontal gyrus (Hickok & Poeppel, 2007). As with the dorsal stream, the ventral auditory processing stream is underpinned by important white matter tracts, with direct stimulation of both the Inferior Frontal-Occipital fasciculus (IFOF; a white matter tract connecting occipital cortex and frontal cortices via the inferior and middle temporal lobes) and the Uncinate Fasciculus (UF; with terminations in the orbital/lateral frontal regions and temporal pole/parahippocampal gyrus) producing consistent semantic paraphasias (Duffau, Gatignol, Moritz-Gasser, & Mandonnet, 2009; Gil-Robles et al., 2013; Harvey, Wei, Ellmore, Hamilton, & Schnur, 2013; Papagno et al., 2011). Additionally, the UF has been shown to have a reduced structural integrity (i.e., lower fractional anisotropy, with higher Mean, Axial and Radial diffusivities) in patients with semantic dementia relative to healthy controls (Agosta et al., 2010). Furthermore, patients with semantic dementia, who have impairments in single word comprehension, are found to consistently have degeneration in the temporal pole region of their brains (Hodges & Patterson, 2007) and research in patients with Primary Progressive Aphasia (PPA) find significant correlations between the degree of atrophy in the left temporal pole and single word comprehension deficits (Mesulam, Thompson, Weintraub, & Rogalski, 2015).

Lastly, the results of a meta-analysis of functional imaging studies that investigated speech perception by DeWitt and Rauschecker (2012) found evidence for a processing hierarchy which followed the theoretical direction of the ventral stream moving progressively along the left superior temporal gyrus/sulcus. Activation specific to phonemes was found in the mid-STG in a region just posterior to the anterior-lateral HG, activation related to words was shown in a more anterior STG region whilst activation to phrase length utterances occurred in an anterior and subjacent region to that activated most strongly by words within the superior temporal sulcus.

In summary, considerable evidence exists for the presence of a functional hierarchy of processing in spoken word comprehension. Within this hierarchy core regions of PAC perform low level acoustic analysis of all sounds irrespective of intelligibility levels. The level of abstract processing then gradually increases with surrounding belt and parabelt regions being responsive to different acoustic structures but are still unconcerned by the level of intelligibility. At the top of the hierarchy are regions both anterior and posterior of PAC which appear to analyse the incoming acoustic signal at an abstract level and are more concerned with intelligibility and linguistic extraction irrespective of difference in surface acoustic structure. These anterior and posterior regions fall into ventral and dorsal streams of processing, with the ventral stream predominantly responsible for mapping sound onto semantic meaning whilst the dorsal stream maps sound onto articulatory motor plans.

Neural Responses to Distorted Speech

As stated above, increasing the intelligibility of a speech signal has repeatedly been associated with increasingly higher activation in anterior and posterior regions of the superior temporal gyrus (Evans, Kyong, Rosen, Golestani, Warren, McGettigan, Mourao-Miranda, et al., 2014; Narain et al., 2003; Scott et al., 2000). In contrast, decreasing the intelligibility of a speech signal by introducing conditions which make successful perception and comprehension of speech harder to achieve are associated with increased activation in the posterior STG. When comparing BOLD signal responses to speech heard in quiet compared to speech at an SNR of either +20dB or -5dB, Wong, Uppunda, Parrish, and Dhar (2008) found an increase in bilateral posterior

STG activation for the speech in noise relative to the clear speech condition. Interestingly, the activation in the left pSTG continued to increase with decreasing intelligibility, i.e., greater activation for -5dB vs +20dB and clear speech, whilst the right pSTG showed increased activation in response to the addition of noise but activation did not vary as a function of the noise level, i.e., activation was equal between the -5dB and +20dB noise conditions. In addition, when using speech stimuli that had been time-compressed to differing degrees, Poldrack et al. (2001) found a convex activation response function in the left pSTG. This study used four different rates of time-compression, with speech compressed to either 60, 45, 30 or 15 percent of its original length. Poldrack et al. (2001) found that activation in the pSTG increased linearly from the 60 percent compression condition, which participants could comprehend well, up to the 30 percent compression rate, in which participants struggled. With activation subsequently decreasing in response to speech presented at 15 percent of its original length, this condition was performed at a rate nominally above chance level by the participants. Similar increases in activation in the pSTG in response to speech that has been distorted in some way have also been found for accented (Adank, Davis, & Hagoort, 2012) and noise-vocoded speech (Scott, Rosen, Lang, & Wise, 2006). The response in this region has been ascribed to a number of different roles. For example, given its response to speech and non-speech acoustic stimuli, it has been suggested that the posterior STG, specifically the Planum Temporale, a region just posterior to the PAC on the superior temporal plane within the lateral sulcus, functions as a computational hub. Within this role as a hub, the PT is believed to act by segregating the incoming signal from the acoustic environment before subsequently matching these signals onto previously learned spectrotemporal representations, this process is referred to as auditory scene analysis. Furthermore, it is hypothesised that the PT is functionally connected to higher order cortical areas that would perform the task of object recognition and high-level perception/comprehension (Griffiths & Warren, 2002). The increased activation of such a computational hub when perceiving speech that has been distorted makes theoretical sense, as the incoming signal would not initially match up to the pre-existing internal acoustic-phonetic templates of the perceiver. As a result, the PT would have to work harder to identify the correct phonemic or word boundaries in the signal to accurately parcellate the incoming signal before attempting to match these segments onto pre-existing representations. Upon initial exposure and/or with increasing distortion, this process

would require an increasing level of computational power to perform, hence the increasing activation, however as auditory perceptual learning occurs with increasing exposure the hub would become more efficient in matching the incoming signal to newly formed/modified internal representations allowing greater comprehension and reduced activation.

Activation related to decreasing intelligibility has also been found in the inferior frontal gyrus (IFG) and ventral premotor regions (PMv). Whilst these regions are implicated in the dorsal stream of functional processing, they are more traditionally associated with speech *production* rather than *perceptual* processes. In addition to pSTG activation, Poldrack et al. (2001) observed activation in three regions of the left posterior inferior frontal gyrus. Each of these three regions followed the same convex pattern of activation as the pSTG with initial increases in activation as speech became harder to comprehend but was still intelligible before reduced activation for speech that was heavily distorted and unintelligible. Similar activation was also observed in the IFG for speech embedded in noise (Adank et al., 2012) and for time-compressed speech in the ventral premotor cortices (Pelle, McMillan, Moore, Grossman, & Wingfield, 2004). Initially this increased activation in the inferior frontal and ventral premotor cortices was hypothesised to be related to increased syntactic processing that also occurs in inferior frontal regions of the brain. However, Hervais-Adelman, Carlyon, Johnsrude, and Davis (2012) also found activation in the ventral premotor cortex when participants heard single words that had been noise-vocoded and Osnes, Hugdahl, and Specht (2011) used fMRI to show that when subjects heard a consonant-vowel syllable that ranged from clearly audible and understandable to overly distorted noise, their premotor cortex was only activated during an intermediate stage where the consonant-vowel syllable became intelligible but was still very distorted. At either end of the continuum where the syllable was either clearly perceivable or complete noise, the premotor cortex showed no significant activation. The increased activation in response to either CV syllables and/or single words precludes the possibility that the increase is syntax processing related. Instead it is suggested that when the incoming signal is ambiguous the motor cortex is “drawn in” to help in comprehension by emphasising that some of the sounds heard are producible by the human articulators and are involved in speech. Then, as the sounds become less ambiguous, the input from the motor cortex is reduced as the temporal lobes are able to comprehend the

sounds with less effort (Tremblay & Small, 2011). Hervais-Adelman et al. (2012) argue that this engagement of the premotor cortex is automatic (and therefore not a process which the individual perceiver chooses to perform) in adverse listening conditions where the purely acoustic form is insufficient to allow successful comprehension.

Evidence for the engagement of non-speech regions during difficult listening conditions have also repeatedly been shown. Most notably, Vaden et al. (2013) found that activation in a network of cingulo-opercular regions of young adult listeners correlated significantly with successful word recognition in noise performance on subsequent trials. This study used multi-talker babble at two different signal-to-noise levels (+3dB SNR and +10dB SNR) and found that the harsher SNR (+3dB) was associated with overall greater levels of activation in the cingulo-opercular network compared to the easier listening condition. In addition, it was found that the extent of the cingulo-opercular activation for one trial was related to the level of success on the following trial, with an elevated level of activation in this network significantly correlated with better overall performance on the following task. This result suggests that activation in this region is related to assisting in task performance/adaptation in difficult listening conditions. However, it is not believed that this region is performing speech-adaptation specific functions. Instead, the cingulo-opercular network has been shown to be activated by tasks that require cognitive control to optimise performance and is activated by all sensory systems, not just auditory/speech related systems. Instead of making speech specific adjustments, Eckert, Teubner-Rhodes, and Vaden (2016) propose that this cingulo-opercular network engages in performance monitoring and increasing cognitive effort with the aim of maintaining a stable level of performance across a task. Reductions in activation of this network on a trial by trial basis are associated with poorer performance on the task following the reduction (Eichele et al., 2008; Weissman, Roberts, Visscher, & Woldorff, 2006) with increased activation in this region related to better performance on the following task (Eckert et al., 2016; Sadaghiani & D'Esposito, 2015; Vaden, Kuchinsky, Ahlstrom, Dubno, & Eckert, 2015; Vaden et al., 2013).

Neural Mechanisms of Auditory Perceptual Adaptation

Whilst previous research has compared the change in neural activation associated with distorted speech to that of clearly intelligible undistorted speech, very few of these studies specifically investigated the neural changes that occur *during* the period of adaptation. Instead most of these studies include a period of training where participants can familiarise themselves with the distortion before the functional imagining occurs (Davis & Johnsrude, 2003). In contrast, Adank and Devlin (2010) were specifically interested in the neural processes that occur during adaptation when listeners are first exposed to the distorted stimuli. In their experiment, participants performed a sentence verification task (e.g. “Cobras crawl around on their bellies” vs “Chairs crawl around on their bellies”) for two types of speech: time-compressed speech (which participants had no previous exposure to) and clear, uncompressed speech produced at a normal tempo. In agreement with the intelligibility research, Adank and Devlin (2010) observed increased activation for the time-compressed (less intelligible) speech in bilateral posterior superior temporal gyrus and the left ventral premotor cortex. However, the pattern of activation within these regions changed as participants adapted to the acoustic manipulation. Activation in the left pSTS and left PMv was significantly larger in response to time-compressed speech compared to clear speech, however activation for the clear speech condition remained at a constant (albeit significantly lower than time-compressed) level throughout. Whereas activation in relation to the time-compressed speech was significantly greater during the first block of the study, representing the period of initial exposure, before gradually declining as participants adapted to the time-compressed stimuli and returned to a level comparable to the clear speech condition by the third block (within 48 sentences). In contrast in the right hemisphere, increased activation was observed in an anterior STG and posterior STS region. In both of these regions activation related to the clear speech decreased monotonically throughout the study in contrast to the left hemisphere regions and in the right posterior STS activation reduced significantly from initial exposure to the time-compressed sentences through to the end of the study. However, in the right aSTG despite an initial reduction in activation after initial exposure to time-compressed speech, activation in this region remained at a significantly elevated level throughout. Adank and Devlin (2010) argue that as adaptation related neural responses occurred in both auditory and motor regions then the process of auditory perceptual

learning and adaptation to distorted speech must involve changing the sensitivity of both auditory and motoric cues. The premotor cortex is traditionally associated with the selection and execution of motor sequences. Therefore, during perception and adaptation to distorted speech stimuli it is hypothesised that the premotor cortex internally simulates the movement that would be required to produce the perceived sounds. This information is then sent along to the auditory cortex in the form of an efference copy to be used to predict the sensory consequences of the modelled motor action. In this way, the novel acoustic patterns perceived during exposure to the distorted speech are mapped onto existing articulatory motor plans. A limitation of these results however relates to the correlational nature of functional imaging data in linking changes in the blood oxygen level in a cortical region with stimulus exposure and task performance. Whilst it is possible that such changes in the BOLD signal are reflective of critical task related activation, such changes in blood oxygen levels could also be epiphenomenal and misleading, thus limiting the strength of conclusions from functional imaging research.

Transcranial Magnetic Stimulation

Transcranial Magnetic Stimulation (TMS) is a neurophysiologic technique that allows for non-invasive stimulation of the human brain through the application of strong but short magnetic pulses that enable us to modulate the underlying neural activity in conscious, healthy human subjects (non-invasively). The principle of TMS is based on Faraday's theory of Electromagnetic induction (1831) which states that a pulse of electric current sent through a wire coil generates a magnetic field and the rate of change of the magnetic field determines the induction of a secondary current in any nearby conductor. In TMS, the pulse is sent along the TMS coil, reaching its peak and returning to zero in less than a millisecond, the very rapid nature of the pulse induces a magnetic field perpendicular to the plane of the coil that also rises/falls rapidly in time. The rapidly fluctuating magnetic field passes unimpeded through the scalp and skull of the participant and induces a current in the brain. If the induced current is of sufficient intensity it will depolarise the neurons in the targeted region and depending on the region and the type of stimulation this can have either an inhibitory or excitatory affect. At a cellular level, it is still largely unknown how TMS works however based

on physiological response timings stimulation is believed to be of the axons as opposed to the cell body with axons most likely to be activated when in a spatially varying field, such as, in the presence of axonal bending, or axonal boundaries e.g. with soma or bouton (Ilmoniemi, Ruohonen, & Karhu, 1999). The effect of a single pulse of TMS lasts 5-30 milliseconds and affects a random percentage of fibres with the population of fibres activated changing depending on the orientation of the coil. The focality of a single pulse of TMS is usually measured on the cortical surface and depends on the type of coil but is estimated to have a surface spatial resolution of 5-20mm (Deng, Lisanby, & Peterchev, 2013). By inducing electrical currents in the brain which modulate and disrupt the ongoing activation within a given region, TMS can be used to demonstrate causality between a cognitive process and a specific brain region and as a result can be used to complement other neuropsychologic techniques (such as fMRI, EEG) which are purely correlation in nature.

Outline of the Thesis

The main aim of the current thesis is to add to our understanding of the neural and cognitive mechanisms affecting perceptual adaptation to distorted speech.

In Chapter 2 I investigate the rate and extent to which individuals are capable of adapting to three different types of speech distortion; time-compressed speech, noise-vocoded speech and speech in noise. In addition, I investigate what, if any, impact individual speaker characteristics has on the adaptation process and further investigate how individual differences in an array of cognitive mechanisms are associated with individual differences in adaptation performance.

In Chapter 3, I will present a series of experiments which aimed to find the most effective TMS protocol to non-invasively impair speech perception in healthy human adults.

In Chapter 4, I will use the TMS protocol deemed to be most effective in Chapter 3 to investigate the functional relevance of the left ventral premotor and left superior temporal cortices in adaptation to time-compressed speech.

Finally, in Chapter 5, the originality and impact of the experimental work presented in this thesis will be considered in the context of previous research and current neurobiological models of speech perception. In addition to considering the ways in which the research contained within this thesis adds to the field as a whole, this final chapter will also consider the limitations of the presented research and suggest possible directions for future research.

All experiments presented in this thesis were approved by the University Research Ethics committee (UCL #0599/001).

Chapter Two

Adaptation to Distorted Speech: Cognitive Mechanisms and the Effect of Listening Distortion and Speaker Variability

Introduction

The majority of everyday communication occurs in the presence of a myriad of distortions that all combine to make speech perception and comprehension challenging. The ability to perceptually adapt to these distortions is essential for successful comprehension. Goldstone (1998) refers to perceptual learning as *“relatively long-lasting changes to an organism’s perceptual system that improves its ability to respond to its environment and are caused by its environment”* (pg.585). With respect to auditory perceptual learning, this would involve a shift in perception whereby an individual is able to demonstrate the ability to either detect, discriminate or identify an acoustic stimulus after a period of exposure despite initial attempts to do so being unsuccessful (Watson, 1980). Perceptual adaptation of this kind has been shown to occur after just a few minutes’ worth of exposure to time-compressed (Fairbanks & Jr., 1957; Mehler et al., 1993; Voor & Miller, 1965) and noise-vocoded speech (Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005; Loizou, Dorman, & Tu, 1999; Shannon et al., 1995) as well as speech embedded in noise (Cainer, James, & Rajan, 2008; Song, Skoe, Banai, & Kraus, 2012). It is hypothesised that this adaptation occurs as a result of a change in the attention-weighting process with listeners shifting their attention from task-irrelevant to task-relevant cues (Adank & Devlin, 2010; Golomb, Peelle, & Wingfield, 2007). In the context of a temporal distortion such as time-compressed speech, changes in attention are believed to occur at a phonological level of processing, whilst for spectral manipulations such as noise-vocoded speech changes in attention and perception are believed to occur at a lexical/semantic level of processing (Davis et al., 2005; Pallier, Sebastian-Gallés, Dupoux, Christophe, & Mehler, 1998).

In the following chapter, two experiments will be presented wherein the rate of adaptation to three different types of speech distortion (time-compressed, noise-vocoded, speech in noise) will be investigated in a single set of participants. Both experiments use stimuli recorded from more than one speaker whilst in experiment

one, individual differences in a battery of audiological and cognitive measures are also assessed to investigate the underlying cognitive mechanisms supporting perceptual adaptation.

Time-Compressed Speech

Time-compressed speech refers to the complete auditory presentation of speech stimuli at a faster than normal rate, for example, a 1000 millisecond token compressed to 50 percent of its original length would subsequently be presented in its entirety in 500 milliseconds. Using this manipulation provides the ability to manipulate stimuli and create speech tokens that when presented appear to have been produced at a rate far quicker than would normally be presented. This faster than normal rate will subsequently put a strain on the auditory perceptual system and thus affords the opportunity to measure the rate of adaptation and make inferences about the processes that underlie the perceptual adjustments. Secondly, the degree of compression is quantifiable and therefore the effects of five percent compression can be compared to those of 10, 20, 50 or 75 percent compression in a systematic fashion. Finally, using modern techniques, time-compression can be applied to the stimuli to manipulate the temporal characteristics of the speech stimuli (i.e., make them shorter or longer) with very minor distortions of the spectral characteristics of the sound (Dupoux & Green, 1997). Such specificity of temporal manipulation without noticeable spectral degradation was not possible when investigations into the effects of time compression were first investigated. For instance, Garvey (1953) found that increasing the level of compression results in lower levels of intelligibility, however the impact of compression rate on subsequent intelligibility does not appear to be linear. Mean intelligibility was above 90 percent correct identification for stimuli shortened up to two and a half times their original length, whilst stimuli compressed to three and a half times their original length results in 50 percent correct identification performance and compression to four times the original length results in performance below 10 percent correct. These results suggest that there is a large amount of redundancy within the system as it appears to be remarkably robust to appreciable levels of distortion with only extreme distortions producing noticeable deficits. Garvey used a “chop-splice” technique where regular segments of the taped recording were physically cut out

before the tape was spliced back together and re-recorded. Note that while this method compresses the speech signal, it also results in part of the word being removed, making it hard to separate the effects of increasing time compression from the effects of partial word deletion. Despite the partial removal of spectral detail, the “chop-splice” method results in speech stimuli with overall equivalent spectral characteristics (e.g. equivalent formant frequencies) to the original stimuli, whereas previous methods of simply accelerating the recording during presentation shifted the frequency content. Garvey (1953) showed that the use of acceleration which results in frequency shifting had a far greater impact on perception than shortening the sound whilst maintaining the frequency content. Identification rates were at less than 10 percent correct for the frequency shifted sounds shortened to two and a half times their original length but at 93 percent correct for the non-frequency shifted stimuli. Whilst modern day techniques do not require the physical cutting up and splicing of tape recordings, the techniques used for time-compression do involve the deletion and overlapping of pitch periods with minimal impact on spectral details and maintenance of the fundamental frequency (Moulines & Charpentier, 1990).

Whilst the foundational work of Garvey (1953) provided numerous insights into the effect that shortening a speech stimuli has on perception and identification, it was another twelve years before any researchers investigated the effect that practice and repeated exposure has on comprehension of time-compressed speech. Initially, Voor and Miller (1965) hypothesised that successive trials of listening to very rapid speech would *not* result in significant improvement in comprehension. However, this hypothesis was rejected when significant improvements in comprehension occurred, furthermore the rate of adaptation to the compressed stimuli occurred in a relatively short period of time. Individual differences in adaptation rate are likely to exist but on average asymptotic performance was observed after approximately seven minutes worth of exposure which was equivalent to roughly 2700 time-compressed words.

Subsequent studies have repeatedly shown that adaptation to time-compressed speech occurs with exposure with recent studies showing that exposure to as few as 20 sentences (200 words) is sufficient for adaptation to occur (Adank & Janse, 2009; Dupoux & Green, 1997; Pallier et al., 1998; Peelle & Wingfield, 2005; Versfeld & Dreschler, 2002). Despite the rapid rate of adaptation Dupoux and Green (1997) found that the extent of adaptation was dependent on the degree of compression. In their

experiment, participants either heard sentences compressed to 45 percent or 38 percent of their original length. Both compression rates resulted in adaptation, however the rate of adaptation was far quicker for the less compressed (45 percent) condition, with adaptation occurring within the first five sentences, compared to the 38 percent condition, where 15 sentences were required to reach asymptotic performance. Additionally, the overall level of adaptation was significantly higher in the less compressed condition with final accuracy levels of nearly 80 percent compared to 40 percent in the more heavily compressed condition. The authors argue that when perceivers encounter speech produced at different rates, they need to retune their perceptual criteria with more time required to retune as the speaking rate increases. This would account for the differential slopes of adaptation in the study with the perceptual adaptation/retuning process taking longer in the more heavily compressed, 38 percent condition.

Dupoux and Green (1997) also investigated whether adaptation to compressed speech transfers to different speakers and whether or not differential intervening rates of compression disrupt the normalisation process. In one experiment, participants heard two sets of five sentences spoken by a male speaker and then a third set of five sentences spoken by a female speaker (or vice versa, half of the participants heard 10 sentences spoken by the female and then five by the male speaker). It was found that the change in speaker overall had little or no impact on listeners' performance, with only a very local impact on performance observed, i.e., performance dropped slightly for the first two sentences after the change in speaker before a rapid recovery in the final three sentences from the new speaker. A similar local effect on performance despite an overall low impact was observed when adaptation to compressed sentences was interrupted by exposure to uncompressed sentences or sentences compressed to a lesser extent (e.g., 50 percent of original duration). Performance again dropped on the two sentences immediately following the sentences of less compressed nature before a rapid recovery. This suggests that adaptation to time-compressed speech occurs in a rapid time frame (less than 20 sentences), is relatively stable once the retuning process is complete and is generalisable from trained stimuli to untrained stimuli/speakers.

Whilst adaptation to time-compressed speech occurs over a very short time window, there is evidence to suggest that further improvements in performance can occur despite initial suggestions that performance asymptotes after 20 sentences.

Peelle and Wingfield (2005) found that in comparison to an older group of adults, younger adults appear to show continued improvement up to and potentially beyond 40 sentences, whilst the older adult's performance appears to asymptote by 20 sentences. Furthermore, Golomb et al. (2007) showed that when participants (both younger and older adults) returned for further testing weekly for three weeks after initial exposure, continual improvements were observed from the first session to subsequent sessions. The largest amount of improvement occurred in the first week, supporting the notion that the majority of adaptation to time-compressed speech occurs in a very short time window, however further improvements are possible with the improvements observed in sessions two to four being smaller in magnitude but consistent. The benefit of multiday training was subsequently investigated by Banai and Lavner (2012) who found that multiday training, involving five practice sessions in between two test sessions spaced 10-14 days apart, resulted in significantly more perceptual learning than was observed in a group of participants who only had the exposure of the pre- and post-test sessions. More importantly, the degree of adaptation was found to be more generalisable in the trained group than the untrained group. When tested on new compressed sentences spoken by the same speaker, both groups showed improvement, however when tested on new sentences spoken by a different male or a female speaker, the trained group performed significantly better than the untrained group (who also showed improvement for the new male speaker but not the female speaker). This is important as it supports the notion that further training/exposure to time-compressed speech can lead to generalisations of learning. However, the degree of generalisation was shown in both groups to be specific, with neither the untrained nor the trained group showing adaptation to time-compressed individual words after exposure to compressed sentences.

This latter finding is an important point as it leads to the question of what adjustments are being made to the perceptual system that allows generalisation of learning across sentence tokens and speakers but not from sentences to individual words. Banai and Lavner (2012) refer to the Reverse Hierarchy Theory (RHT) of perceptual learning from Ahissar, Nahum, Nelken, and Hochstein (2009) and argue that their results suggest that there are two stages in the process of adapting to time-compressed speech. According to RHT initial attention weighting is by default allocated to higher level representations, whilst subsequent exposure and training leads

to a shift in focus from higher-level representations to a more fine-grained low level, pre-lexical representations, the authors posit suprasegmental and sound envelope information to be potential candidates of low level characteristics that receive greater attention with further training. The higher-level focus of the preliminary adaptation period could explain why an initial level of generalisation occurs during the early exposure and training phase, i.e., although the trained group performed at a significantly higher level, both groups performed better on sentences spoken by a different male after initial testing. Yet, as stated neither group was able to transfer this learning to the individual word tokens, which Banai and Lavner (2012) argue is due to the shift in attentional focus from higher levels to more fine-grained stimuli-specific information. This view of shifts in attention is supported by Golomb et al. (2007) who also argue that adaptation to time-compressed speech requires the shifting of attention from task irrelevant cues to cues more relevant for the perception of time-compressed speech. Golomb et al. (2007) argue that the normal speech rate contains phonemes that are prototypical and match predefined internal representations, however when first exposed, the time-compressed speech tokens are too far away from these prototypes to be correctly identified. The role then of training is to redefine and retune the boundaries of our internal phonemic prototypes such that they incorporate the extremes encountered in time-compressed speech.

This notion of adaptation to time-compressed speech consisting of a function of shifts in sensitivities/attention at the pre-lexical level is supported by a series of studies which investigated the transfer of perceptual learning of time-compressed speech between languages. Pallier et al. (1998) conducted a series of experiments that investigated whether higher level, lexical/semantic, processes are essential for adaptation to time-compressed speech. To do this they presented participants with time-compressed sentences in either a familiar or unfamiliar language and observed the impact of this exposure on a series of test sentences in the participants' native language. Most interestingly, when Spanish monoglots were presented with training/habituation sentences in Catalan, a language of which they had no comprehension, performance on subsequent test sentences in their native Spanish was equivalent to a group of participants that had been trained in Spanish, with both groups performing significantly better than a control group who had no exposure to any practice sentences prior to testing. Furthermore, when English monoglots were

exposed to time-compressed sentences in either French, a syllable-timed language (i.e., where each syllable has the same length) that is very different to English or Dutch which also has many differences to English but shares the characteristic of being a stress-timed language (i.e., where stressed syllables occur at regular intervals but intervening syllables vary in length), performance was improved for English time-compressed test sentences following exposure to the Dutch sentences but not the French training stimuli. Pallier et al. (1998) argue that these results show that semantic understanding and lexical access are not essential for adaptation to time-compressed speech to take place. Additionally, the finding that certain pairs of languages show transfer of learning whereas others do not would suggest that adaptation to time-compressed speech requires more than shifts in attention at the level of raw acoustic properties. If this was the case then exposure to time-compressed sentences in nearly any language would boost performance on subsequent testing. Instead, given that English and Dutch are rhythmically similar, Pallier et al. (1998) hypothesise that adaptations must be occurring at a pre-lexical linguistic level, presumably at the level of phonology. In support of this hypothesis, Sebastián-Gallés, Dupoux, Costa, and Mehler (2000) also presented practice sentences to native Spanish speaking in either Italian or French, both of which are Romance languages as is Spanish, and therefore most of the lexical roots of all three languages derive from Latin. Alternatively, participants heard practice sentences in either English or Japanese, neither of which are Romance languages and therefore have very different lexicons, morphological systems and syntax to Spanish (French and Italian). The results show that participants who were exposed to Italian sentences during practice performed at an equivalent level to those trained on compressed Spanish (native) sentences, and significantly better than participants who were trained on all other languages. This shows that when the foreign language shares many properties with the speakers' native one, subjects benefit from previous exposure to the compressed speech and can transfer this exposure to their native language even in absence of semantic understanding. In order to rule out the possibility that the benefit provided by exposure to Italian sentences was not related to the shared lexical roots of the two languages, Sebastián-Gallés et al. (2000) conducted a follow-up experiment where participants were presented with training sentences in Greek, a non-Romance, syllable-timed language which has very limited lexical overlap with Spanish. Sebastián-Gallés et al. (2000) found that when Greek was used as the practice stimuli, participants performed equally as well as when

Spanish was used, and both groups performed significantly better than control subjects who received no exposure to time-compressed speech prior to testing. These results are consistent with the notion that adaptation to time-compressed speech occurs at a level more abstract than at the level of raw acoustic properties. If it existed at the raw acoustical level then the transfer of learning from male to female speakers or only from rhythmically similar languages would not be expected. Instead, given all of the evidence it appears that adaptation to time-compressed speech most likely occurs as a result of shifts in attentional focus at the phonological level.

An important point to make however is that the studies discussed thus far have all used speech that has been artificially time-compressed. This form of distortion has been used as it stresses the perceptual system enough to allow investigation of the process of perceptual adaptation. However, research suggests that responses to artificially time-compressed speech, which is largely only encountered in the laboratory and responses to naturally fast speech, encountered on a regular basis in real world situations are different. In the creation of artificially time-compressed speech all elements of the speech signal are compressed in a linear and equal fashion resulting in a literally shortened version of the original signal. However, when we naturally produce fast speech, not all parts of the speech signal are affected equally, for example, consonant durations are generally less affected than vowel durations and some segments can be entirely deleted (Adank & Janse, 2009). As a result, naturally fast speech involves greater spectrotemporal deviations from normal speech than artificial time-compressed speech. Such differences in the creation/production of the different speech formats impacts upon the perception. Adank and Janse (2009) investigated the degree to which people adapt to artificially and naturally time-compressed sentences and whether learning of one form would generalise to and assist with subsequent perception of the other form. Overall performance on the natural-fast sentences was significantly poorer than for the normal rate sentences whilst no difference was observed between the time-compressed and normal rate sentences. The lack of significant difference between the artificially compressed and normal sentences could be due to the fact the time compression of the artificially time-compressed sentences was matched to the compression of the natural-fast sentences. The significant difference therefore between the naturally fast condition and the artificially fast and normal conditions is likely due to the added difficulty imposed by the syllable

elisions and deletions found mainly in production of vowels in naturally fast speech compared to the artificially compressed condition which contains very minor spectral changes in the signal. Therefore, whilst the artificial time-compressed sentences were shorter in duration compared to the normal sentences, the amount of compression (on average 46 percent of original length) and the lack of significant spectral manipulation likely contributed to the lack of a significant difference between the artificially time-compressed and normal sentences. Despite the overall poorer performance for the naturally fast speech, participants were still able to adapt to this condition but adaptation took longer to occur with approximately 30 sentences required before greater levels of comprehension occurred. Additionally, and of most interest, this study found a transfer of learning effect from artificially compressed speech to naturally compressed speech, with performance starting at a higher level for natural fast speech if preceded by artificially time-compressed speech, but the reverse pattern was not observed, i.e., learning did not transfer from naturally fast speech to artificially time-compressed speech. Adank and Janse (2009) argue, with reference to RHT (Ahissar et al., 2009), that this difference in transfer of learning is due to the fact that the artificially time-compressed sentences posed less of a challenge to the perceptual system and therefore participants were able to process this stimuli at a higher level. Consequently, when subsequently faced with the harder, naturally fast speech tokens, participants were better able to move past the higher-level features and focus attention on lower-level cues. This fits with the prediction of RHT that transfer of learning occurs when an easy condition is followed by a more difficult condition. The lack of transfer from the naturally fast to the artificially fast sentences is due to the fact that in order to comprehend the naturally fast sentence, participants had to immediately focus attention to lower level properties of the signal, thus resulting in learning of more stimulus specific information that does not transfer as well to alternative stimuli.

In summary, time-compressed speech refers to a form of speech that has been artificially shortened with very minimal impact on spectral content. The rate and amount of adaptation for this form of distortion is dependent on the amount of compression, however in general the vast majority of adaptation occurs rapidly and within the first 10-20 sentences of exposure. Based on the transfer of learning from one language to another and from one speaker to another, it is believed that adaptation

to this form of distortion does not require comprehension but instead is most likely to occur as a result of shifting attention at the level of phonological representations.

Noise-Vocoded Speech

Noise-vocoded speech is created by dividing a speech signal up into a predefined number of logarithmically spaced frequency bands/channels. A smoothed amplitude envelope for each independent channel is then extracted and used to modulate band limited noise. The shaped noise across all of the bands is then recombined creating a speech signal that contains enough detail to be intelligible but with significantly less spectral detail than the original signal (Davis et al., 2005; Shannon et al., 1995). In this way, noise-vocoded speech provides an opposite form of distortion to time-compressed speech with manipulation of the spectral detail but preservation of the temporal structure.

In noise-vocoded speech, the level of intelligibility is associated with the number of channels initially used to divide up the speech signal, with more channels resulting in subsequent higher levels of intelligibility. Loizou et al. (1999) assessed the number of channels required to produce a high level of comprehension, and found that there was no statistically significant difference in performance when the speech was vocoded to between eight and 16 channels with performance on a single word identification task above 90 percent in all cases. Use of five or six channels impaired performance relative to the eight and above conditions but performance overall for these conditions was still around 90 percent correct. However, below five channels recognition performance drops rapidly as the vocoded speech becomes very difficult to comprehend with performance dropping to 63 percent for four channels and below 30 percent for two channels. In a pre- and post- training study, Huyck and Johnsrude (2012) found that the only group to significantly benefit from training were the group that attended to noise-vocoded speech. Two other groups that heard the noise-vocoded speech but were instructed to attend to either a visual or auditory distractor using a target detection task did not show any benefit of exposure. This finding suggests that attention to noise-vocoded speech is necessary for effective perceptual learning.

Unlike the rapid adaptation that occurs in time-compressed speech, adaptation to noise-vocoded speech requires a longer time window with Davis et al. (2005)

observing a steady linear increase in word report scores across 30 sentences that had been vocoded into six channels. Across the 30 sentences, participants appeared to improve from roughly 20 percent of words correctly reported to roughly 60 percent however no obvious asymptotic performance is apparent which strongly suggests that further adaptation was possible. The most significant contribution of the Davis et al. (2005) study however is one which suggests that unlike time-compressed speech, adaptation to noise-vocoded speech appears to be somewhat dependent on higher level lexical information. The intelligibility of noise-vocoded speech is heavily influenced by prior knowledge of sentence content, e.g. comprehension of a vocoded sentence without any prior context will be low, however if you hear the same sentence in a clear non-vocoded form immediately before hearing the vocoded form, the vocoded sentence will appear dramatically more intelligible than before. This concept is referred to as pop-out, as the content of the vocoded sentence following the clear version suddenly “pops out” of the otherwise distorted stimuli. Across a series of experiments, Davis et al. (2005) asked participants to report words from a vocoded sentence. Once the participant had finished reporting the words they had perceived, they either heard or read the clear version of the previously vocoded sentence followed by another repetition of the vocoded sentence. It is important to note here that the exposure to the clear sentence came after the participant had responded to the initial vocoded sentence, therefore any benefit of exposure to the content via the clear sentence would only be evident on different sentences, depending on whether or not training on vocoded sentences could be generalised to untrained stimuli. Davis et al. (2005) found performance to be significantly improved in both groups that were exposed to the clear content of the sentence, either aurally or visually, before a second repetition of the vocoded sentence, compared to a group who only heard the vocoded sentences without an intervening clear sentence. The enhanced performance observed after exposure to the clear forms supports the notion that being aware of the content enables greater levels of adaptation to noise-vocoded speech. In addition, the equal benefit gained from the clear auditory and clear written forms of the sentence suggests that information supporting adaptation to noise-vocoded speech must be non-acoustic and therefore at the phonological level or above.

This hypothesis is supported by Hervais-Adelman, Davis, Johnsrude, Taylor, and Carlyon (2011) who investigated whether adaptation to speech that had been

noise-vocoded within a very specific frequency range would transfer to speech that had been vocoded in a different and non-overlapping frequency range. Participants were trained on either low-passband (50-1406Hz) or high-passband (1934-5000Hz) noise-vocoded speech. It was found that learning bandpass filtered speech transferred between regions with participants that switched from low passband to high passband stimuli (or vice versa) after 20 sentences performing at a level equivalent to participants that been trained entirely on the high passband sentences. Hervais-Adelman et al. (2011) argue that the transfer of learning from one frequency region to another suggests that the observed adaptation related modifications must be occurring at a level of internal perceptual representations that are not dependent on frequencies. Given the tonotopic organisation of the auditory pathways/primary auditory cortex, these results support the notion that perceptual learning for noise-vocoded speech is not assisted by acoustic structures and most likely occurs beyond the primary auditory cortex.

To further investigate the level at which adaptation predominantly occurs, Davis et al. (2005) investigated the effect of training participants on 20 vocoded non-word sentences. Participants were trained, via passive listening, on 20 real or 20 non-word vocoded English sentences and then were tested on 20 real vocoded English words. Overall, the group trained on the real words performed significantly better than the group trained on non-words, additionally, the group trained on non-words performed at a level that was equivalent to subjects that were completely naïve, i.e., subjects that had no prior history of exposure to vocoded speech. This result suggests that adaptation to noise-vocoded speech is dependent on either lexical, semantic and/or syntactic information with phonological information being less important, i.e., the exact opposite of time-compressed speech.

In a further experiment, Davis et al. (2005) trained different groups of participants on vocoded sentences with different syntactic structures: normal prose, i.e., real word, syntactically and semantically correct English sentences; syntactic prose, i.e., real word, syntactically correct sentences with no semantic meaning; jabberwocky prose, i.e., real function words combined with content non-words, maintaining syntactic integrity but without semantic meaning and finally non-word sentences which lack both syntactic and semantic integrity. Training on normal and syntactic vocoded prose resulted in significantly better performance for real word

vocoded sentences than any of the other training conditions, with jabberwocky significantly better than non-word prose. From this it can be concluded that a top-down, lexically driven mechanism is involved in perceptual learning of noise-vocoded speech, with information at the lexical level being used to make alterations to perceptual processes at a pre-lexical level.

However, this conclusion has been debated, in a follow up to the previous study Hervais-Adelman, Davis, Johnsrude, and Carlyon (2008) investigated the effect of feedback order (i.e., distorted [respond] clear-distorted vs distorted [respond] distorted-clear) and lexicality for individual noise-vocoded words as opposed to sentences. Word report scores were noticeably lower for individual vocoded words compared to whole sentences (60 percent correct for sentences compared to 39 percent for isolated words). However, this study replicated the results of Davis et al. (2005) and found that word report accuracy was significantly higher if in the order of post response feedback, the clear word was presented before the vocoded word was repeated. Hervais-Adelman et al. (2008) argue that the clear before distorted feedback order provides superior benefits in learning noise-vocoded speech because the clear form provides the learner with a “teaching signal” against which the participant can compare the following distorted repetition, this teaching signal is believed to help the auditory perceptual system to accurately map distorted sounds onto internal representations, supporting the notion that higher level information must be present for effective learning. The validity of this “teaching signal” is evident in a study by Loebach, Pisoni, and Svirsky (2010), who found a significant effect of training when participants received a written form of the heard sentence at the same time as the vocoded sentence was repeated (replicating Davis et al., 2005). Loebach et al. (2010) however found no significant effect when participants simply heard a clear form of the vocoded sentence without a repetition of the vocoded sentence. In this instance, post-training performance was equivalent to naïve participants who received no training in noise-vocoded speech comprehension (replicating Davis et al. 2005, distorted [respond] distorted-clear feedback paradigm). Whilst this training paradigm provides a teaching signal, the participants have to compare this signal to a previously heard vocoded sentence as opposed to the same sentence in vocoded form following presentation of the clear signal, this order appears to be less effective for perceptual learning.

In contrary to Davis et al. (2005) however, Hervais-Adelman et al. (2008) found that training with individual noise-vocoded non-words was equally as useful as a tool for learning as were vocoded individual real words, this directly contradicts the results of Davis et al. (2005) who found no benefit of using non-word sentences during training. Hervais-Adelman et al. (2008) observe that this discrepancy could be due to storage limitations in short term memory which make it harder to hold an entire sentence of vocoded non-words, whilst storage of a single vocoded non-word places considerably less strain on phonological STM. As a result individual vocoded non-words can more easily be retained for comparison to internal representations and used to assist perceptual retuning. However, when Davis et al. (2005) presented noise-vocoded non-word sentences to participants, they also provided one group with a written version of the sentence, this was designed to specifically relieve the strain placed on phonological STM when trying to remember an entire non-word vocoded sentence. Despite this assistance, no benefit of training with non-word vocoded sentences was found. Overall, these results suggest that if lexical information is available then it provides significant contributions in the process of retuning the perceptual system and constraining the set of expected phonemic strings. However, if explicit lexical information is unavailable, limited adaptation can occur for shorter stimuli based purely on phonological information.

In summary, noise-vocoded speech refers to a form of speech that has been spectrally degraded with preservation of temporal structure by the recombination of amplitude modulated bandpass noise channels. The rate and amount of adaptation is predominantly dependent on the number of channels with more channels resulting in overall higher intelligibility. Two significant differences exist between adaptation to time-compressed sentences and noise-vocoded sentences; firstly, adaptation to noise-vocoded sentences can be observed within 20 sentences but occurs gradually and does not appear to reach asymptote until much later than time-compressed sentences. Secondly, whilst the specific balance between lexical and phonological levels of processing is still yet to be fully established, research thus far supports the notion that lexical information appears to be far more important in noise-vocoded adaptation than for time-compressed adaptation where lexical-semantic information does not appear to be required for adaptation to occur.

Speech in Noise

Unlike time-compressed and noise-vocoded speech, speech in noise does not require any manipulation of the speech signal itself but instead involves embedding the signal in a form of background noise, such as white noise, speech-shaped noise or multi-talker babble, thus representing an environmental distortion as defined by Mattys et al. (2012). As the vast majority of everyday conversations occur in the presence of background noise such as competing conversation, music or traffic, understanding the processes behind adaptation to speech in noise has important real-world implications. Most studies investigating speech in noise perception use a staircase procedure whereby the signal to noise ratio between the speech and background noise starts at either a favourable or unfavourable level and is varied dependent on participant performance. As a result, establishing rate of adaptation for speech in noise is different from time-compressed or noise-vocoded speech where the levels of distortion are usually set at a specific level and held constant across participants. Using such a calibration procedure Peelle and Wingfield (2005) found the SNR at which participants were able to correctly report 30 percent of heard words. Once their SNR had been established participants then heard 20 sentences at this level to investigate if adaptation to speech in noise occurs. Peelle and Wingfield (2005) found no evidence of improvement/adaptation across the 20 sentences in either a younger or older adult population. Whilst this initially suggests that adaptation to speech in noise does not occur, it is more likely that participants were not provided with enough sentences in which to adapt given the difficulty of the task, i.e., performance started at 30 percent correct. When investigating adaptation rates for speech presented at a level that produced 50 percent accuracy in recall, Cainer et al. (2008) found that within 15 sentences, participants appear to reach an initial level of asymptotic performance, but with more testing performance can improve further until roughly 90 sentences where improvement levels again appear to asymptote. Thus, suggesting that adaptation to speech in noise is possible but the rate of adaptation, as with time-compressed and noise-vocoded speech is dependent on the level of distortion.

Support for the added benefits of long-term auditory training was shown by Zaballos, Plasencia, Gonzalez, de Miguel, and Macias (2016) who compared the ability to perceive speech at three different signal-to-noise ratios in a group of experts: experienced air traffic controllers, who are constantly exposed to noisy radio

communications, and a control group: normal hearing individuals. Disyllabic Spanish words were presented to both groups of participants at one of three signal-to-noise ratios (+5, 0, and -5dB). The expert group performed significantly better than the control group in all three signal-to-noise ratio conditions with the largest group differences in the most difficult condition (-5dB) supporting the notion that with extensive training individuals can show greater adaptation to speech in noise. Burk, Humes, Amos, and Strauser (2006) further investigated the effect of training on perception of speech in noise with specific focus on the degree of generalisation. Participants were trained in seven, 60 minute, sessions over a two-week period on a set of individual monosyllabic words spoken by the same speaker, they were then subsequently tested on a subset of the trained words by the same speaker as well as untrained words by the same speaker, to investigate generalisation across items, as well as being tested on the trained words when spoken by a new set of speakers. Evidence of generalisation of learning from one speaker to another would be indicative of learning affecting the lexical representation of the words themselves as opposed to the listeners adapting to cues specific to the speaker they heard during training (in addition, participants received orthographic feedback during training when they answered incorrectly). Burk et al. (2006) found that the rate of improvement declined (i.e., learning slowed down over time) from 5.9 percent in the first hour, to 4.3 percent in the second to less than half a percent in the remaining hours of training suggesting that adaptation had reached asymptotic levels after two hours of training. Additionally, performance improved significantly post-training for both the trained and the untrained words, suggesting that training for speech in noise can generalise to new items (although the magnitude of improvement was noticeably smaller for the untrained words relative to the trained words). Furthermore, no significant difference between speakers was found for the trained words and only one difference was found for the untrained words, where participants actually performed significantly better for a new speaker than for the speaker they were trained with. This pattern of results makes a theory of perceptual adjustments based on speaker specific characteristics hard to support and instead suggests that adaptation to speech in noise occurs as a result of adjustments to pre-existing lexical representations or memorisation of new ones. In a follow-up experiment, Burk et al. (2006) re-tested a group of participants six months after initial training and found residual benefits of training, with trained word identification being significantly greater than the pre-train test scores of six months

earlier. Although participants were below the level that they were at the immediately post-training time point, it appeared that they had retained about half of their initial improvements. In agreement with the findings of Zaballos et al. (2016) this suggests that training to perceive speech in noise is beneficial in both the short and longer term. Additionally, in the first experiment Burk et al. (2006) provided orthographic feedback to participants during training on trials where they responded incorrectly. In a subsequent experiment, this feedback was removed and it was found that overall improvements were smaller than observed in the first experiment and the learning did not generalise from trained to untrained items. Burk et al. (2006) conclude that repeated exposure to speech embedded in noise is sufficient to effect improvements in perceptual performance, however the size of the learning effect is significantly enhanced when lexical information is available, similar to noise-vocoded speech. Bradlow and Alexander (2007) and Mayo, Florentine, and Buus (1997) both found an effect of lexical predictability when native (and non-native) listeners of English were asked to report sentences heard in noise, with highly predictable sentences being perceived with greater accuracy than sentences low in predictability. This supports the notion that perception of speech in noise is aided by lexical-semantic context effects which presumably enable constraining the set of internal predictions related to words the listener may expect to hear.

In summary, speech in noise refers to the distortion of a speech signal by embedding it in a level of background noise. Adaptation to this form of distortion is dependent on the severity of the signal to noise ratio, with more severe SNRs requiring longer time windows for adaptation to occur. As with time-compressed and noise-vocoded speech, evidence of adaptation to speech in noise can be observed after as few as 15 sentences worth of exposure, however in order to reach asymptotic levels of performance, numerous hours (~two) worth of training are required. Finally, in analogy with noise-vocoded speech, adaptation to speech in noise appears possible with simple exposure but for the greatest amount of adaptation to occur, access to and retuning of internal lexical items is required.

Transfer of Learning from one Distortion to Another

There is one area of surprising scarcity in the adaptation literature, that is nearly all of the existing adaptation studies have investigated the effects of adaption in only one source of distortion (Bradlow & Bent, 2008; Cainer et al., 2008; C. M. Clarke & Garrett, 2004; Davis et al., 2005; Hervais-Adelman et al., 2008; Mehler et al., 1993; Pallier et al., 1998; Zaballos et al., 2016). Where adaptation has been investigated in multiple forms of distortion, studies have either used different groups of participants or have not reported the (transfer of) learning effects (Davis & Johnsrude, 2003; Peelle & Wingfield, 2005). Thus far only two studies have specifically investigated whether learning/adaptation to one form of distortion transfers to the learning of another distortion (Bent, Baese-Berk, Borrie, & McKee, 2016; Borrie, Baese-Berk, Engen, & Bent, 2017). Bent et al. (2016) investigated the recognition of words in phrases across three speech varieties; a nonnative accent, a regional dialect and ataxic dysarthric speech within the same group of participants to assess whether listeners that attain a level of proficiency in one form of distortion obtain the same level of proficiency across all other distortions. Results show a significant correlation between performances in the nonnative accented speech condition and both the regional dialect and dysarthric speech conditions, suggesting that individuals who were able to successfully perceive speech in the nonnative condition were also more successful in both of the other conditions. However, no correlation was found between performance in the regional dialect condition and the dysarthric speech condition. The authors conclude that these results suggest that listeners are not “globally skilled” at perceiving speech that deviates from expected norms. Instead, it appears that different individuals possess a capacity to map the acoustic-phonetic eccentricities found in certain types of distortions (but not all) onto words in their mental lexicons. However, in a follow up study, Borrie et al. (2017) investigated the overlap in ability to correctly report words spoken by an individual with dysarthria or words presented in noise. The authors found a significant positive correlation between performances in the two conditions and concluded that these results suggest similar cognitive-perceptual processes are used to aid comprehension in both conditions, i.e., it appears that participants do possess a global skill that allows them to adapt to a relatively equal level when the speech signal is distorted in an array of forms.

As discussed previously, despite the vast amount of research that has been conducted to investigate perceptual adaptation to distorted speech, only Bent et al. (2016) and Borrie et al. (2017) have explicitly investigated how individuals deal with and adapt to different types of distortions. Therefore, one of the main aims of the research in this chapter is to further investigate how individuals adapt to multiple distortions. Specifically, whether individuals who show an ability to adapt to speech that has been temporally manipulated (i.e., time-compressed) can also adapt equivalently to speech that has been spectrally manipulated (i.e., noise-vocoded speech) and/or speech that has been affected by environmental distortion, i.e., speech in noise. Participants in the experiments described in this chapter were predicted to adapt to each of the individual conditions in the same way as outlined above, i.e., rapid adaptation to time-compressed speech and slower and less extensive adaptation to the noise-vocoded and time-compressed conditions. However, it is not known whether performance in one condition will equate to a relatively similar performance in all conditions. Due to the similar dependence on accessing and retuning existing internal lexical representations it is possible that individuals who are capable of adapting to noise-vocoded speech may show similar adaptation to speech in noise (or vice versa). However, lexical access has been shown to be redundant when adapting to time-compressed speech, this reliance on different perceptual adjustments may result in no transfer of learning from this condition to the other conditions or vice-versa. As a result, participants could fall into two groups, each of which may use different adaptation strategies, those that are better able to employ higher-level lexical strategies to adaptation and show learning for both noise-vocoded speech and speech in noise and a second group who are better able to employ lower-level phonological strategies and thus adapt better to the time-compressed speech. Alternatively, as Bent et al. (2016) suggest, participants may possess (or indeed lack) a “global skill” for adaptation irrespective of the form of manipulation (temporal, spectral or environmental) and thus performance across all three manipulations will be related.

Transfer of Learning from one Speaker to Another

In the same way that the vast majority of the speech adaptation literature has only investigated how listeners adapt to a single form of distortion, the majority of studies

in the field have also only used stimuli based on recordings from a single speaker during the adaptation phase (Adank & Janse, 2010; Davis et al., 2005; Hervais-Adelman et al., 2008; Janse & Adank, 2012; Mehler et al., 1993; Pallier et al., 1998). The assumption of using one speaker is that the adaptation that occurs is distortion specific and not speaker specific thus once adaptation has occurred for one speaker this perceptual learning will transfer to other speakers. This has generally been found to be accurate with across speaker generalisation found for both time-compressed speech (Dupoux & Green, 1997) and speech in noise (Cainer et al., 2008). Furthermore, adaptation to foreign accented speech, especially non-native phoneme contrast training has been shown to benefit from a high-variability approach. For example, when investigating adaptation to foreign accented speech, Bradlow and Bent (2008) found that exposure to multiple Chinese-accented speakers of English during an adaptation phase led to a level of adaptation in American participants that was equivalent to training with the specific Chinese-accented speaker who the participants were tested on. Interestingly, the effect of training with a speaker that was different to the test speaker resulted in significantly poorer adaptation than training with multiple speakers or training only with the test-specific speaker. The high efficacy of multiple speaker training has been shown for successful learning of the /r/-/l/ contrast in English by Japanese listeners (Shinohara & Iverson, 2018) and in assisting English listeners to learn Chinese lexical tone contrasts (Wang, Spence, Jongman, & Sereno, 1999). Bradlow and Bent (2008) conclude that “...*exposure to high-variability training stimuli promotes, rather than interferes with, perceptual learning for speech...*” (pg.722).

Therefore, in addition to investigating the degree of learning in the same participants across a range of distortions the experiments presented in this chapter also investigated whether adaptation to time-compressed speech, noise-vocoded speech and speech in noise was enhanced or impaired by the inclusion of multiple speakers during adaptation. If adaptation is distortion-specific, then participants will perform equally well when adapting to a single speaker as when adapting with multiple speakers. However, if the perceptual learning involves a level of adaptation to the idiosyncrasies of each individual speaker then inclusion of multiple talkers could have one of two effects; (1) participants benefit from the rich source of variation within each distortion and this assists overall adaptation, resulting in better performance than

would be observed from single-talker training. (2) Constantly switching from one speaker to another inhibits adaptation to speaker-specific characteristics and result in overall poorer adaptation than would be observed for adaptation to a single speaker (where no switching is required). Based on the findings from the second language learning and foreign accented speech adaptation literature, it is expected that adaptation will be distortion specific and the effect of training with multiple speakers will be equivalent to that observed when trained with a single speaker.

Individual Differences in Perceptual Adaptation

The ability to adapt to distorted speech has been related to a range of cognitive factors, yet no comprehensive model currently exists that explains which factors are most important and how these factors interact with the type of adverse condition. Thus far, the association between four audiological and cognitive abilities and subsequent adaptation to distorted speech have been investigated most: individual hearing thresholds; working memory; selective attention/inhibition and vocabulary knowledge (see Table 1). The impact of individual differences in hearing ability has predominantly been investigated in older populations where difficulty perceiving speech especially in the presence of background noise is a common trait. Whilst overall hearing thresholds have been found to be associated with poorer overall performance on distorted speech tasks (Adank & Janse, 2010; Akeroyd, 2008; Janse & Adank, 2012), deficits in speech in noise perception persist even when the older participants have healthy hearing (Tun, 1998; Tun & Wingfield, 1999; Wong et al., 2009); or when statistical (Bilodeau-Mercure, Lortie, Sato, Guitton, & Tremblay, 2015) or experimental (B. C. Moore, Peters, & Stone, 1999) adjustments are made to counteract the increased auditory thresholds. This suggests therefore that it is not just the decline of the auditory periphery that is causing the speech in noise deficit, effective listening appears to also rely upon general cognitive processes (Golomb et al., 2007).

In addition to hearing thresholds, working memory has also been associated with individual differences in perceptual adaptation to distorted speech. The Ease of Language Understanding (ELU) model (Rönnberg et al., 2013; Rönnberg, Rudner, Foo, & Lunner, 2008) emphasises the role of working memory capacity specifically in suboptimal conditions where the incoming perceived signal is distorted and does not

match any internal phonological representations. In this situation, working memory is required to initially keep track of the incoming signal before subsequently assisting in inferring meaning from the incomplete information gained from the distorted incoming signal. In support of this model, working memory has been shown to be an important cognitive mechanism when perceiving speech in noisy environments. Akeroyd (2008) suggests that after hearing thresholds, working memory is the most effective cognitive mechanism in explaining individual differences in performance on tasks requiring perception of speech in noise. This conclusion is in agreement with research investigating perception of an unfamiliar accent (Banks, Gowen, Munro, & Adank, 2015; Janse & Adank, 2012), which found an association between unfamiliar accent perception and working memory in both younger and older participants. However, no relationship was found between individual working memory capabilities and performance requiring perception of foreign-accented (Gordon-Salant, Yeni-Komshian, Fitzgibbons, Cohen, & Waldroup, 2013); frequency compressed (Ellis & Munro, 2013) noise-vocoded (Erb, Henry, Eisner, & Obleser, 2012; Neger, Rietveld, & Janse, 2014), or speech in an array of noise backgrounds (Boebinger et al., 2015). With a meta-analysis from Füllgrabe and Rosen (2016) concluding that for young listeners who are assessed to have healthy hearing thresholds, individual differences in working memory accounts for less than two percent of the variance in speech in noise perception.

A similar inconclusive relationship has also been found between individual differences in attention switching/inhibition. Huyck and Johnsrude (2012) found that attending specifically to noise-vocoded speech led to overall higher levels of adaptation in their participants. Additionally, attention switching or inhibition has also been linked with greater overall performance for foreign (Janse & Adank, 2012; Tao & Taft, 2017) and novel accented speech (Adank & Janse, 2010; Banks et al., 2015) with a mediating effect in the perception of noise-vocoded speech (Erb et al., 2012). However, Bent et al. (2016) found no relationship with foreign accented or dialect accented speech, Ellis and Munro (2013) found no relationship with frequency compressed speech and Boebinger et al. (2015) found no relationship between attention switching/inhibition and speech in noise. Finally, whilst individual differences in vocabulary knowledge have mainly only been investigated in studies of adaptation to accented speech, the results thus far have been more consistent, with

greater vocabulary knowledge associated with greater adaptation to accented speech across numerous studies (Adank & Janse, 2010; Banks et al., 2015; Bent et al., 2016; Janse & Adank, 2012; Neger et al., 2014). The inability to establish conclusive relationships between individual differences in audiological and cognitive measures and perceptual learning of degraded speech thus far is most likely due to the use of different test batteries and different distorted speech conditions across experiments. The final aim therefore of this research chapter is to establish the extent to which individual differences in a single battery of audiological and cognitive assessments are associated with performance across three different types of speech distortion in the same set of participants, with particular focus on the degree of overlap or divergence in how each cognitive measure relates to each separate speech condition.

Table 1 - Summary of the previous research investigating the link between distorted speech and cognitive mechanisms.

Primary Author(s)	Year	Form of Speech Distortion	Cognitive Mechanism Investigated	Specific Test	Relationship Found	Reported Statistic	
Adank & Janse	2010	Novel Accent	Hearing	PTA	Yes	$\beta = 0.07$ $SE =$ $0.03, p <$ 0.05	
			Attention Switching	Trail Making Test	Yes	$\beta = 0.92$ $SE =$ $0.42, p <$ 0.05	
			Information Processing	Digit-Symbol Substitution Test	No		
Erb et al	2012	Noise-Vocoded	Working Memory	Forward Digit Span	No		
				Backward Digit Span	No		
				Non-word Repetition	No		
			Psychoacoustic	Amplitude Modulation	Yes	$r = -$ $0.51, p <$ 0.05	
Janse & Adank	2012	Foreign Accent	Hearing Thresholds	PTA	Yes	$\beta = -$ $0.04 SE$ $= 0.01,$ $p <$ 0.001	
				Auditory Short term memory	Auditory non-word repetition	Yes	$\beta = 0.06$ $SE=0.02,$ $p < 0.01$
				Working Memory	WAIS Digit Span Task	Yes	$\beta = 0.02,$ $SE =$ $0.01, p <$ 0.05
				Vocabulary Knowledge		Yes	$\beta = 0.49$ $SE=0.17,$ $p < 0.01$
				Selective Attention	Flanker Task	Yes	$\beta = -$ $2.37 SE$ $= 0.71, p$ < 0.001
				Attention Switching	Trail Making Test	No	
Ellis & Munro	2013	Frequency Compressed in noise	Working Memory	Reading Span Test	No		

			Attention Switching	Trail Making Test	No	
Gordon-Salant et al	2013	Accented speech in noise	Hearing Thresholds	PTA	Yes	$p < 0.01$
			Information Processing	WAIS-III Digit Symbol	No	
				WAIS-III Digit Search	Not Tested	
			Working Memory	WAIS-III Digit Span	No	
				WAIS-III Letter-Number Sequencing	Not Tested	
*Neger et al	2014	Noise-Vocoded	Hearing Thresholds	PTA	No	
			Working Memory	WAIS Backward Digit Span	No	
			Information Processing	Digit Symbol Substitution	No	
			Attention Switching	Trail Making Test	No	
			Statistical Learning	Symbol version of Artificial Grammar Learning Test	Yes	$\beta = -9.22$ $SE = 4.05$, $p = 0.023$
			Vocabulary Knowledge		Yes	$\beta = 1.48$ $SE = 0.56$, $p = 0.009$
Banks et al	2015	Unfamiliar Accent	Inhibition	Stroop Test	Yes	$\beta = 0.29$, $p = 0.004$
			Vocabulary Knowledge	WASI vocabulary subtest	Yes	$\beta = -0.24$, $p = 0.02$
			Working Memory	Reading Span Test	Yes (indirect)	$\beta = 0.09$, $p < 0.01$
Bent et al	2016	Regional, non-native and disordered Accent	Selective Attention	Flanker Test	Yes (indirect via vocab.)	$\beta = 0.18$, $p < 0.001$
			Cognitive Flexibility	Intra Dimensional Set Shift Test	Yes (indirect via vocab.)	$\beta = 0.33$, $p < 0.001$

				Extra Dimensional Set Shift Test	Yes (indirect via vocab.)	$\beta = 0.32$, $p < 0.001$
			Vocabulary Knowledge	Peabody Picture Vocabulary Test	Yes	$\beta = 0.076$, $p < 0.001$
Boebinger et al	2016	Speech in noise	Psychoacoustic	Frequency discrimination	No	
				Duration discrimination	No	
			Working Memory	WAIS Forward Digit Span	No	
				WAIS Backward Digit Span	No	
			Non-verbal IQ	WASI Matrix Reasoning subtest	Yes	$r^2 = 0.146$, $p < 0.05$
			Inhibition	Stroop	No	
			Selective Attention	Trail Making Test	No	
Tao et al	2017	Foreign Accent	Attention Switching	Trail Making Test	Yes	$B = 3.11$, $t = 2.81$, $p = 0.006$
			Working Memory	WAIS-IV Reading Span Test	No	
			Inhibition	Stroop	Yes	$B = 6.11$, $t = 3.12$, $p = 0.002$
			Information Processing	WAIS-IV Coding Subtest	Yes	$B = -3.58$, $t = -2.86$, $p = 0.005$
				Alphabet Backwards Test	No	

Notes: * study found different relationship between cognitive measures in younger and older groups of participants. Only the effects in the younger group are reported here.

Summary of research aims

1. Determine the extent to which learning of one type of speech distortion transfers to the learning of other forms of distortion.
2. Through the use of multiple speakers, determine the extent to which the vocal idiosyncrasies of the different speakers either promotes or inhibits perceptual learning of distorted speech.
3. Determine the extent to which individual differences in a battery of audiological and cognitive assessments are related to performance in each of the different speech conditions and whether the pattern of associations is consistent across the different conditions.

Experiment One

Methods

Participants

Ninety participants (mean age 21years 3months \pm 2.74; range 18-30; 65 females) were recruited for this experiment. All participants were native British English speakers, had normal or corrected to normal vision, were right-handed as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971; mean score 80.16 ± 20.07 ; range 25-100) and were highly educated (mean years of education 15years 10months \pm 1.67; range 13-20). No participants reported a history of speech, language, neurological or psychiatric disorder. All participants gave written informed consent and were compensated with monetary payment or course credit.

Procedure

Participants were tested on a battery of audiological and cognitive assessments in addition to the main speech adaptation task. All testing was performed in a double-walled soundproof room and lasted up to 90 minutes. Half of the participants were tested on the audiological/cognitive measures followed by the adaptation task; the other half of participants had the opposite order to avoid fatigue related effects on

results. No significant effect of procedural order was found ($p>0.25$) and therefore in all subsequent analyses the data is collapsed across this variable.

Audiological Assessments

Two audiological assessments were performed: (1) Pure Tone Audiometry (PTA) using a clinical audiometer (Maico, MA 41) with each ear tested separately at octave frequencies between 250 and 8000Hz. For each participant, a pure tone average (average threshold across all measured frequencies) was computed for both the left and right ear. (2) Speech Recognition Thresholds (SRT) were used to assess the lowest level at which participants could comprehend 50 percent of an aurally presented sentence (Plomp & Mimpen, 1979c). Each test started at +20dB and varied systematically thereafter. Each sentence had five key words, if participants repeated three or more of the key words then the SNR value would decrease on the subsequent trial, initially in steps of -10dB and subsequently in steps of -2dB, thus making the following trials harder to perceive. The SNR value decreased until participants were only able to comprehend two or fewer of the key words at which point the SNR value would initially increase in steps of +6dB and subsequently in steps of +2dB. The first six lists of the IEEE Harvard Sentences (IEEE, 1969) were used (60 sentences; see Appendix A). On average 36 trials/sentences were required to establish each individual SRT. Sentences were presented in the same order to all participants.

Cognitive Assessments

Working Memory was assessed using a forward digit span task. Participants initially heard a set of three numbers and were asked to repeat them back in the same order as heard. This was repeated for six lists. If participants correctly recalled five or six lists correctly, then the list size increased to four numbers and so on until more than one list was incorrectly recalled for a list size. At this point the last correctly recalled list size was taken as the individuals' working memory threshold (see Appendix B).

Vocabulary Knowledge was assessed using the spot-the-word section of the Speed and Capacity of Language Processing (SCOLP) test (Baddeley, Emslie, & Nimmo-Smith,

1993; Baddley, Emslie, & Nimmo-Smith, 1992). Participants were presented with 60 pairs of letter strings (see Appendix C) and were asked to indicate which one of the letter strings per pair spells a real British English word (by circling the correct answer). Reported scores are number of correct identifications out of 60 (see Appendix D).

Attention-Switching was assessed using the trail-making test (TMT; see Appendix E). This task consists of two parts, in part A, participants have to draw a line to connect 25 numbers in ascending numerical order (1-2-3-4 etc.) as quickly as possible. In part B, participants have to draw a line to connect 24 circles; 12 of which contain numbers, and 12 of which contain letters of the alphabet, in an alternating numerical and alphabetic sequence (1-A-2-B-3-C etc.) again the participants were required to do this as quickly as possible. A ratio score between the two parts was the main outcome statistic (part B/part A).

Pattern/Rule Analysis was assessed using the Wisconsin Card Sorting test. In this test participants are required to sort a deck of 128 cards into stacks depending on how they correspond to one of four reference cards. Each card (playing and reference) contains a symbol of a certain shape, colour and size. The participant has to sort the cards depending on one of these features. Critically, participants are initially unaware of how the playing cards and reference cards correspond, with the researcher simply informing them whether each placement is correct or incorrect. After 10 correct placements (for example matching 10 playing cards in front of the corresponding colour matched reference card) the correspondence rule changes and participants must first notice the rule has changed and then find the new rule. Each of the correspondence rules are repeated twice per test (making six rules), the outcome measure reported here is the number of trials required to complete all rules (i.e., two sets of 10 correct placements). Perfect performance would be completing this task in 60 trials.

General Cognitive Ability was assessed using the Montreal Cognitive Assessment (MoCA). Through a one page, 30 point test, the MoCA is able to provide a general overview of cognitive ability with seven main subsections: Visuospatial; Naming;

Attention; Language; Abstraction; Delayed Recall and Orientation (Nasreddine et al. (2005) ; see Appendix F).

Speech Adaptation Task

The main speech adaptation task was a computerised version of the SCOLP speed of comprehension sentence verification test. Participants listened to simple sentences (see Appendix G) in each of the four conditions outlined below and had to decide whether the sentence was true or false, indicating their response by pressing either the left (true) or right (false) key of a standard PC keyboard. All sentences were clearly true (*'Admirals are people'*) or false (*'Admirals have fins'*). Accuracy and response times (RTs) were recorded per trial with adaptation to each condition adjudged via improvements in speed and accuracy of sentence verification.

Stimuli: The auditory sentences were recordings of 192 SCOLP sentences, 96 true and 96 false, with 48 sentences presented per speech condition (clear, time-compressed, noise-vocoded and speech in noise). Sentences varied from three to eight words (mean length 4.44 ± 1.19) with an average of 7.7 syllables per sentence (range 3-18) and an average length of 1.38 seconds (SD 0.28; range 0.82-2.29 seconds). All sentences were recorded by four different male speakers of standard British English. At time of recording all speakers were between 30-32 years of age; and all were born, raised and educated to undergraduate level in South East England (see Table 2 for acoustic information relevant to each speaker). Different sentences from each speaker were used 12 times per condition with the order of speaker randomised. All sentences were saved to separate files with the beginning and end trimmed to zero crossings as closely as possible to the onset/offset of the initial/final speech sounds; resampled to 22050 Hz; peak normalized to 99 percent of maximum amplitude and scaled to 70dB SPL using Praat (Boersma & Weenink, 2011). Stimulus presentation was performed using a custom-made MATLAB 2014a program (The MathWorks Inc., Natick, MA, 2000) and Sennheiser headphones, with all stimuli delivered at a comfortable listening level (preset at 74dB SPL but where necessary this output level was adjusted to fit individual participant preference).

Table 2 - Acoustic Characteristics of the different speakers.

Speaker	Mean f_0	Median f_0	Min f_0	Max f_0	Jitter	Shimmer	Clear Rate (syl/sec)	Compressed Rate (syl/sec)
One	154 (21)	145	93 (24)	287 (126)	0.033 (0.01)	0.093 (0.024)	4.07 (0.77)	10.17 (1.93)
Two	131 (37)	117	83 (6)	239 (155)	0.025 (0.007)	0.09 (0.025)	4.64 (0.91)	11.59 (2.28)
Three	122 (33)	111	74 (4)	238 (150)	0.028 (0.008)	0.103 (0.024)	5.00 (1.04)	12.50 (2.61)
Four	118 (56)	93	70 (5)	230 (173)	0.036 (0.01)	0.115 (0.035)	4.94 (1.02)	12.36 (2.56)

Notes. Standard deviation displayed in parentheses.

f_0 = Fundamental Frequency, Clear and Compressed Rate refer to the number of syllables per second, calculated by dividing the number syllables per sentence by the total length of sentence (including pauses).

Participants' ability to adapt to different types of speech was tested using four different conditions. (1) Time-compressed sentences shortened to 40 percent of their original length, resulting in an average syllable rate of 14 syllables per second (clear speech: 5.57 syl/sec), using the PSOLA algorithm implemented in Praat software (Moulines & Charpentier, 1990). This algorithm initially segments the speech signal into consecutive pitch periods. Adjacent pitch periods are then averaged and smoothed depending on the compression rate producing a speech signal with fewer pitch periods than the original and thus being shorter and appearing compressed (Dupoux & Green, 1997). (2) Noise-vocoded sentences were filtered into four logarithmically spaced frequency bands from 50 to 5000Hz (50-528; 528-1248; 1248-2541; 2541-5000Hz). (3) Speech in noise sentences were embedded in a stream of speech-shaped noise at a signal to noise ratio of -4dB. The spectrum of the speech-shaped noise was derived from the 192 sentences used in the adaptation task. (4) Clear sentences were presented without any manipulation (beyond the zero trimmings, peak normalization etc. outlined above). The clear speech condition was always the first condition participants heard. This was to ensure that any task practice effects were overcome before exposure to the distorted stimuli. Theoretically therefore any improvement in task performance for the time-compressed, noise-vocoded and speech in noise conditions came from the participants adapting to the specific manipulation and was not due to greater familiarity with the task. Order of presentation of the three distorted speech conditions was fully randomised between participants.

Data Analysis

The data were analysed with generalised linear mixed-effects models in MATLAB (R2014b; The Mathworks Inc., Natick, Massachusetts, USA) with the `fitglme` function. The generalised linear mixed-effects model included participant as a random factor and condition (clear, time-compressed, noise-vocoded, speech in noise) and speaker (one-four) as categorical fixed factors and trial number (1-48) as a continuous fixed factor. For the speech condition factor, the clear condition was mapped onto the intercept and speaker one was mapped onto the intercept in the speaker variable. Due to the binary nature of the accuracy data (incorrect = 0 and correct = 1), a binomial logit linking function was included (Jaeger, 2008) between responses and predictor variables. For the response time data a normal identity link function was used between the data (correct responses only) and predictor variables (Ng & Cribbie, 2017). The best fitting model for each data set was established through stepwise model comparisons using likelihood ratio tests. Where no significant difference was found between models, the simplest model will be described. Figure 1 shows the change in accuracy and response times over the course of each condition. The 48 sentences in each condition were divided up into four consecutive blocks of 12 sentences in order to depict this adaptation.

Results

Accuracy

Overall accuracy was highest in the clear speech condition (mean = 95.79 ± 20.09), followed by the time-compressed (mean = 85.37 ± 35.34), speech in noise (mean = 70.09 ± 45.79) and finally the noise-vocoded condition (mean = 60.65 ± 48.85). The results of the generalised linear mixed model show an overall significant effect of trial number ($\beta = 0.001$, $SE_{\beta} = 0.0006$, $p = 0.04$) indicating that performance improved over the trials. Furthermore, the results show an overall effect of condition, with accuracy in the time-compressed ($\beta = -0.117$, $SE_{\beta} = 0.022$, $p < 0.001$), noise-vocoded ($\beta = -0.464$, $SE_{\beta} = 0.025$, $p < 0.001$) and speech in noise ($\beta = -0.317$, $SE_{\beta} = 0.024$, $p < 0.001$) conditions all significantly poorer than the clear speech condition. Finally, a significant effect of speaker was found with perception of the fourth speaker being significantly impaired relative to the first speaker ($\beta = -0.066$, $SE_{\beta} = 0.025$, $p = 0.008$). No

significant interaction effects were found in the model and no significant effects of trial number were found when conditions were analysed separately, indicating that on the whole participants became more accurate as they progressed through the experiment, but the accuracy of individual conditions did not improve significantly (see Appendix H for results of follow-up contrasts).

Response Times

Response times were analysed for correct responses only. RTs were measured relative to the end of each sentence and it was therefore possible for a participant to obtain a negative response time which would represent a correct response made during sentence presentation (as opposed to a positive response time which would represent a correct response after sentence presentation). Overall participants were quickest to respond in the clear speech condition (mean = 409 ± 361), followed by speech in noise (mean = 611 ± 455), then noise-vocoded (743 ± 458) and finally the slowest overall RTs were in response to the time-compressed speech (818 ± 435).

When analysing the four conditions in separate generalised linear mixed models, a significant effect of trial number was found for all conditions: clear ($\beta = -4.78$, $SE_{\beta} = 0.63$, $p < 0.001$); time-compressed ($\beta = -3.73$, $SE_{\beta} = 0.45$, $p < 0.001$); noise-vocoded ($\beta = -2.57$, $SE_{\beta} = 0.608$, $p < 0.001$) and speech in noise ($\beta = -1.66$, $SE_{\beta} = 0.54$, $p = 0.002$). In all conditions participants became quicker to make a correct response as the number of trials increased. When including all conditions in a single model the effect of trial number remained significant ($\beta = -4.683$, $SE_{\beta} = 0.786$, $p < 0.001$). In combination with the increasing accuracy, decreasing response times is indicative of adaptation.

Furthermore, a significant effect of condition was found with participants taking significantly longer to respond in the time-compressed ($\beta = 336.27$, $SE_{\beta} = 36.60$, $p < 0.001$); noise-vocoded ($\beta = 312.04$, $SE_{\beta} = 40.14$, $p < 0.001$) and speech in noise ($\beta = 155.39$, $SE_{\beta} = 37.93$, $p < 0.001$) relative to the clear speech condition which was mapped onto the intercept. Lastly, a significant effect of speaker was also found with response times to speaker three ($\beta = 65.238$, $SE_{\beta} = 31.08$, $p = 0.035$) and speaker four ($\beta = 99.82$, $SE_{\beta} = 37.97$, $p = 0.008$) being significantly slower than speaker one.

The two-way interaction between trial number and speaker was also significant, as the number of trials increased responses to speaker two became slower relative to speaker one ($\beta = 4.754$, $SE_{\beta} = 1.19$, $p < 0.001$), there was no significant effect found for the other two speakers across trial number.

Lastly, the three-way interaction between condition, speaker and trial number was also significant, when analysing the four speech conditions separately, speakers three ($\beta = 66.52$, $SE_{\beta} = 24.92$, $p = 0.007$) and four ($\beta = 94.56$, $SE_{\beta} = 30.6$, $p = 0.002$) produced significantly slower response times than speaker one in the clear condition. There was no other significant speaker by trial number interactions in the other three conditions (see Appendix H for results of follow-up contrasts).

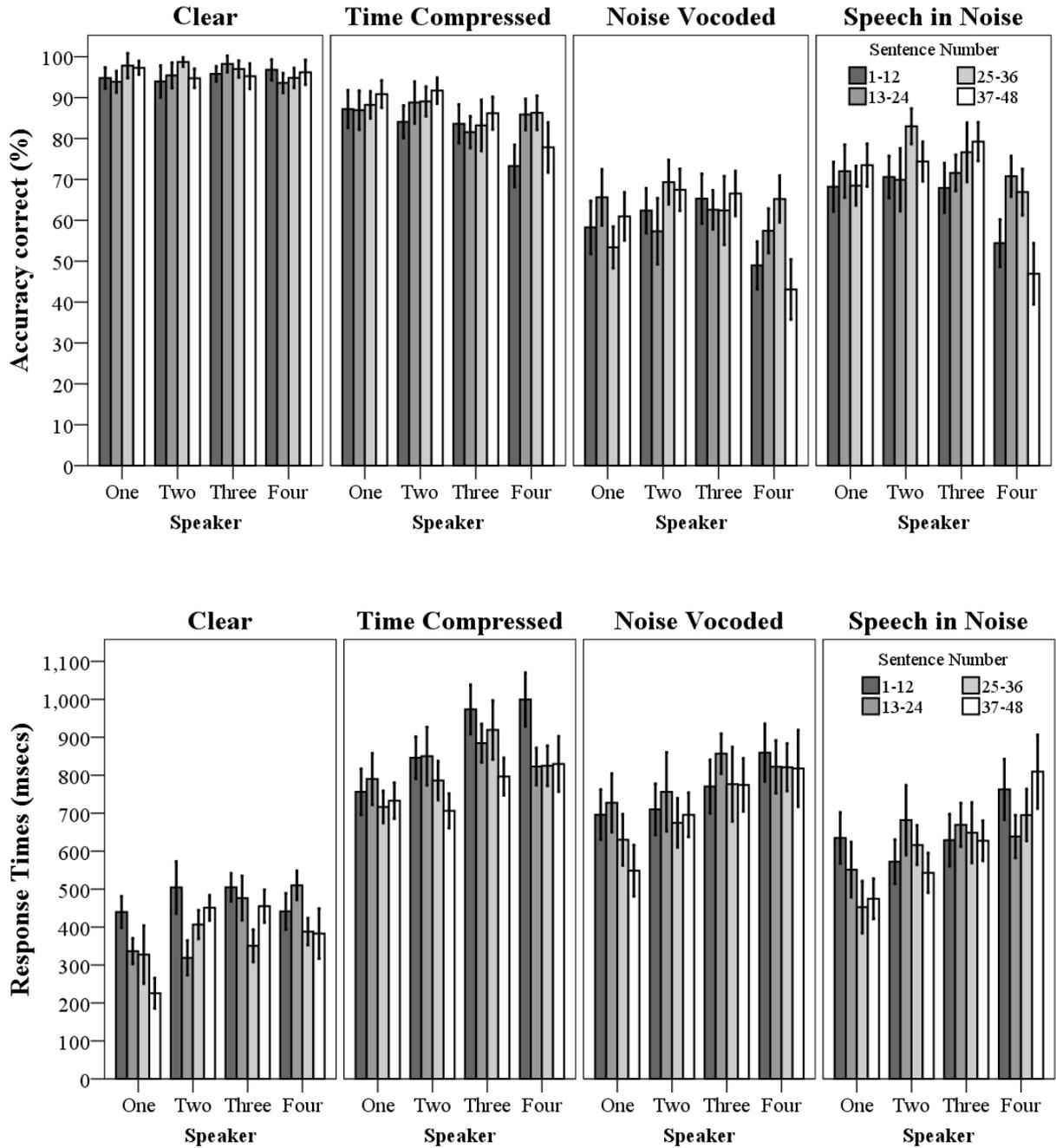


Figure 1 - Accuracy and Response Times for Experiment One across Condition, Speaker and Sentence Number.

Correlation in Performance Between Conditions

Correlations between sentence verification accuracy/response times for each of the four conditions were conducted to establish whether individual participants performed consistently across the different conditions. That is, do participants possess (or lack) a global skill to adapt to all conditions or do individual participants possess a specific ability to adapt to specific distortions (e.g., noise-vocoded speech) but no other distortions. Results for both accuracy and response time data reveal significant positive and monotonic Spearman's rank-order correlations between all conditions (Bonferroni corrected alpha-level $(0.05/6 = 0.008)$; all p values were below this corrected level; see Table 3 and Table 4; Figure 2 and Figure 3). This result suggests that participants do in fact possess (or lack) a general ability to adapt relatively equally in different adverse listening conditions, irrespective of the type (spectral, temporal or environmental) of distortion.

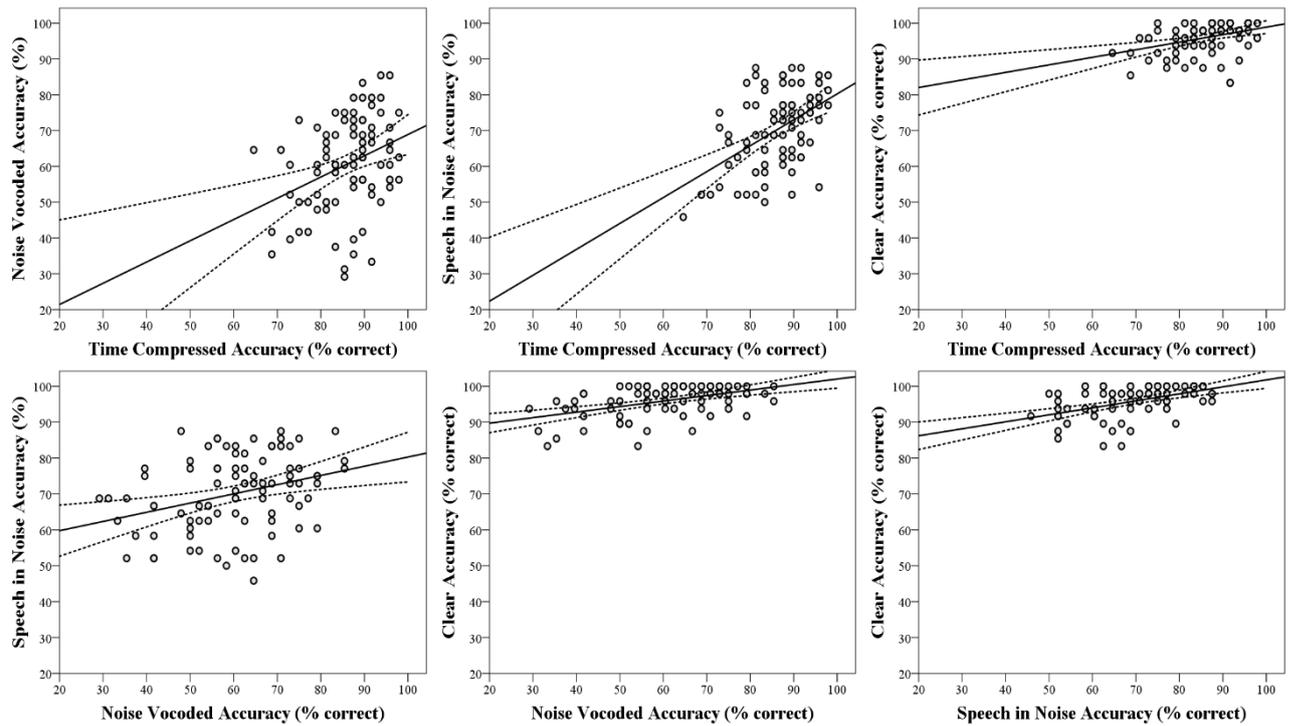


Figure 2 - Scatterplots displaying the relationship between accuracy of performance in each of the adverse listening conditions. Solid line represents the line of best fit. Dotted lines represent 95 percent confidence intervals.

Table 3 - Statistics from individual Spearman's rank-order correlations on accuracy data.

Condition A	Condition B	Correlation coefficient (ρ)	p -value
Time-Compressed	Noise-Vocoded	0.329	0.002
Time-Compressed	Speech in Noise	0.445	<0.001
Time-Compressed	Clear	0.397	<0.001
Noise-Vocoded	Speech in Noise	0.301	0.004
Noise-Vocoded	Clear	0.415	<0.001
Speech in Noise	Clear	0.492	<0.001

Notes. Bonferroni corrected alpha-level = $(0.05/6) = 0.008$

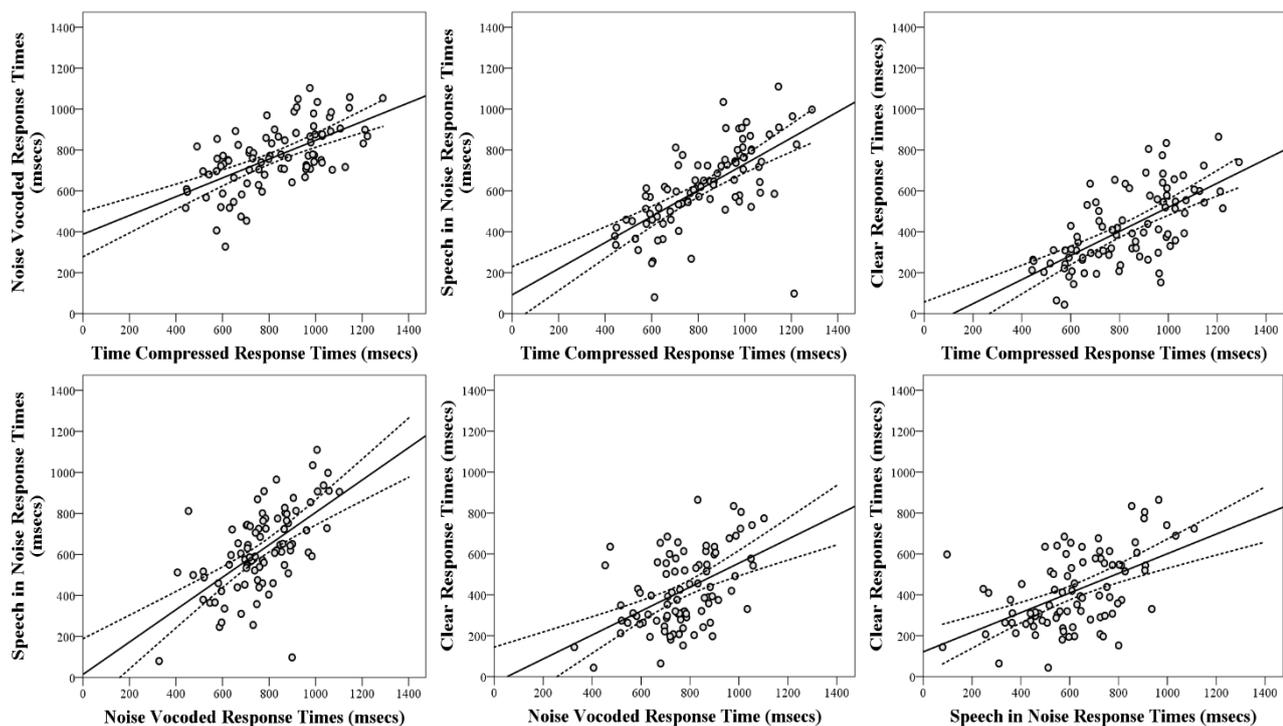


Figure 3 - Scatterplots displaying the relationship between response times performance in each of the adverse listening conditions. Solid line represents the line of best fit. Dotted lines represent 95 percent confidence intervals.

Table 4 - Statistics from individual Spearman's rank-order correlations on response data.

Condition A	Condition B	Correlation coefficient (<i>rho</i>)	<i>p</i> -value
Time-Compressed	Noise-Vocoded	0.607	<0.001
Time-Compressed	Speech in Noise	0.702	<0.001
Time-Compressed	Clear	0.654	<0.001
Noise-Vocoded	Speech in Noise	0.609	<0.001
Noise-Vocoded	Clear	0.451	<0.001
Speech in Noise	Clear	0.499	<0.001

Notes. Bonferroni corrected alpha-level = (0.05/6) = 0.008

Analysis of Speaker Acoustic Characteristics

To investigate whether the different speakers varied significantly in their individual acoustic characteristics, a series of independent Kruskal-Wallis H tests were run for the median fundamental frequency data and syllable rate statistics. Median average f_0 values were statistically significant between speakers, $X^2(3) = 278.54$, $p < 0.001$. Follow-up Bonferroni corrected Mann-Whitney U tests (adjusted alpha = $0.05 / 6 = 0.008$) showed a significant difference in median average f_0 values between all speakers (see Table 5 for statistics). Speaker one had the highest median average f_0 value (145) followed by speaker two (117), then three (111) and finally speaker four had the lowest median average f_0 value (93).

Table 5 - Bonferroni corrected Mann-Whitney U pairwise comparison tests of Median average f_0 values.

Speaker A	Speaker B	<i>U</i> -statistic	z-score	<i>P</i>
One	Two	5817	-11.6	<0.008
One	Three	4850	-12.48	<0.008
One	Four	6559	-10.91	<0.008
Two	Three	10581	-7.219	<0.008
Two	Four	8117	-9.48	<0.008
Three	Four	8791	-8.86	<0.008

Notes. Bonferroni corrected alpha-level = $(0.05/6) = 0.008$

Median syllable rate values were also significantly different between speakers, $X^2(3) = 46.93$, $p < 0.001$. Follow-up Bonferroni corrected Mann-Whitney U tests (adjusted alpha = $0.05 / 6 = 0.008$) showed a significant difference in median syllable rate between speaker one and the other three speakers, no significant difference was found between speakers two, three, and four after correcting for multiple comparisons (see Table 6). Speaker one produced the fewest syllables per second (4.38) followed by speaker two (4.9), then four (5.27) and finally speaker three had the quickest median syllable rate (5.31).

Table 6 - Bonferroni corrected Mann-Whitney U pairwise comparison tests of median syllable rate.

Speaker A	Speaker B	U-statistic	z-score	p
One	Two	13974	-4.09	<0.008
One	Three	12061	-5.85	<0.008
One	Four	12175	-5.73	<0.008
Two	Three	15908	-2.32	0.02
Two	Four	16199	-2.05	0.04
Three	Four	18175	-0.236	0.81

Notes. Bonferroni corrected alpha-level = (0.05/6) = 0.008

Relationship Between Cognitive Assessments and Performance in Different Speech Conditions

Overall performance on each of the audiological and cognitive assessments was high, as would be expected from a homogenous young, highly educated population of participants (see Table 7 for descriptive statistics of performance).

Table 7 - Descriptive statistics for the battery of audiological and cognitive assessments.

Assessment	N	Mean	Std. Dev.	Min.	Max.
PTA Left Ear	90	3.93	4.70	-5.00	25.00
PTA Right Ear	90	5.29	5.55	-2.50	35.00
Speech Reception Threshold	89	-3.80	1.11	-7.00	0.00
Working Memory (Forward Digit Span)	90	6.30	1.21	4.00	9.00
Vocabulary Knowledge	90	50.08	3.75	42.00	58.00
Attention Switching (Trail Making)	90	2.14	0.70	1.01	5.65
Pattern Analysis (Wisconsin)	90	76.94	11.55	63.00	129.00
General Cognitive Ability (MoCA)	89	28.80	1.38	24.00	30.00

Notes. For PTA and Speech Reception Threshold lower numbers equal better performance. Maximum score for: Working Memory task was a list size of 10; Vocabulary Knowledge was 60; Pattern Analysis was 60; General Cognitive Ability was 30.

Prior to establishing the relationship between individual differences in audiological/cognitive ability and performance on each of the speech distortions a Principal Component Analysis (PCA) was run on the audiological and cognitive assessment scores to establish correlations between variables. Initially, average pure tone audiometry thresholds for both left and right ears; speech reception threshold; working memory; vocabulary knowledge; attention-switching (trail-making test ratio); pattern analysis (number of trials required to complete all rules on Wisconsin card

sorting) and general cognitive ability score as assessed via the Montreal Cognitive Assessment were included in the analysis. Scores on the SRT test however were found to have no correlation coefficients above 0.3, indicative of its lack of correlation with other audiological and cognitive measures and was therefore removed, the PCA was re-run on the remaining variables. Bartlett’s test of sphericity was found to be significant indicating sufficient between variables correlations in the remaining measures to be suitable for a PCA ($X^2(21)=71.01, p<0.001$).

PCA revealed three components (see Table 8 below for the rotated loadings matrix) that had eigenvalues greater than one and which, when combined, explained 62.84 percent of the total variance. A Varimax orthogonal rotation was employed to aid interpretability. Component one loads most strongly onto vocabulary knowledge, general cognitive ability and working memory and reflects a form of verbal intelligence. Component two reflects general hearing ability loading most strongly onto the pure tone audiometry thresholds of the two ears. Component three reflects a form of fluid intelligence with loadings onto attention-switching and pattern analysis.

Table 8 - Rotated Structure Matrix for PCA with Varimax Rotation.

Items	Rotated Component Coefficients		
	1	2	3
Vocabulary Knowledge	.827	-.133	-.060
General Cognitive Ability (MoCA)	.762	-.057	-.047
Working Memory	.711	.057	-.048
PTA Left Ear	.039	.816	-.126
PTA Right Ear	-.073	.782	.072
Attention-Switching	.003	-.120	.922
Pattern Analysis	-.264	.397	.445

Notes: Rotation Method: Varimax with Kaiser Normalization. Major loadings for each item have been highlighted in bold.

Multiple linear regressions were conducted to establish if performance in each of the four speech conditions could be predicted based on individual differences in each of the three components and the SRT task (separate analyses were conducted for accuracy and response time data).

The multiple linear regression model significantly predicted overall accuracy performance in the Clear, Time-Compressed and Speech in Noise conditions (all p 's < 0.05; see Table 9 for summary of multiple regression analyses) but not in the Noise-

Vocoded condition ($p=0.17$). In all three significant models, component one from the PCA was the only significant predictor, suggesting a link between higher level verbal intelligence (vocabulary knowledge, working memory and general cognition) and adaptation in these conditions.

For the response time data, the multiple linear regression model was only significant in the Time-Compressed condition $F(4,81)=2.69$, $p=0.037$, adj. $R^2=0.074$ (see Table 10 for summary of multiple regression analyses). Performance in none of the other conditions was significantly predicted (all p 's >0.05). The only significant predictor in the time-compressed model was the second component ($B = 48.93$, $SE_B = 21.20$, $\beta = 0.24$, $p = 0.024$) from the PCA, which most strongly loaded onto differences in auditory threshold.

Table 9 - Summary of Multiple Regression Analyses for Accuracy Data across conditions.

Clear Condition: $F(4,81)=5.32, p=0.001, \text{adj. } R^2=0.208$				
Variable	B	SE_B	β	p
Intercept	92.37	1.57		<0.001
Component One	1.26	0.44	0.29	0.005
Component Two	-0.36	0.43	-0.08	0.41
Component Three	-0.35	0.42	-0.08	0.40
SRT	-0.94	0.40	-0.24	0.02

Time-Compressed Condition: $F(4,81)=6.59, p<0.001, \text{adj. } R^2=0.209$				
Variable	B	SE_B	β	p
Intercept	81.87	2.62		<0.001
Component One	2.86	0.73	0.39	<0.001
Component Two	-0.77	0.73	-0.10	0.29
Component Three	0.71	0.70	0.99	0.31
SRT	-1.06	0.67	-0.16	0.11

Noise-Vocoded Condition: $F(4,81)=1.62, p=0.17, \text{adj. } R^2=0.028$				
Variable	B	SE_B	β	p
Intercept	52.21	5.08		<0.001
Component One	1.87	1.42	0.14	0.19
Component Two	0.39	1.41	0.03	0.78
Component Three	0.53	1.36	0.04	0.69
SRT	-2.14	1.30	-0.18	0.10

Speech in Noise Condition: $F(4,81)=5.08, p=0.001, \text{adj. } R^2=0.161$				
Variable	B	SE_B	β	p
Intercept	67.58	3.96		<0.001
Component One	4.18	1.10	0.39	<0.001
Component Two	-0.81	1.10	-0.07	0.46
Component Three	1.42	1.06	0.13	0.18
SRT	-0.65	1.01	-0.06	0.52

Notes: B = standardised regression coefficient; SE_B = Standard error of the coefficient; β = standardised coefficient. All p -values highlighted in bold are significant at an alpha level of <0.05

Table 10 - Summary of Multiple Regression Analyses for Response Time data across conditions.

Clear Condition: $F(4,81)=0.987, p=0.41, \text{adj. } R^2=0.046$				
Variable	B	SE_B	β	p
Intercept	451.78	74.31		<0.001
Component One	-26.706	20.76	-0.145	0.202
Component Two	23.404	19.98	0.128	0.245
Component Three	-7.712	20.14	-0.042	0.703
SRT	7.799	18.764	0.047	0.679
Time-Compressed Condition: $F(4,81)=2.69, p=0.037, \text{adj. } R^2=0.074$				
Variable	B	SE_B	β	p
Intercept	925.84	78.74		<0.001
Component One	-33.46	21.99	-0.165	0.13
Component Two	46.92	21.18	0.233	0.03
Component Three	1.91	21.34	0.009	0.929
SRT	24.63	19.88	0.135	0.219
Noise-Vocoded Condition: $F(4,81)=0.593, p=0.669, \text{adj. } R^2=-0.02$				
Variable	B	SE_B	β	p
Intercept	842.707	63.14		<0.001
Component One	14.33	17.63	0.093	0.419
Component Two	7.92	16.98	0.051	0.642
Component Three	9.84	17.11	0.063	0.567
SRT	19.38	15.94	0.139	0.228
Speech in Noise Condition: $F(4,81)=0.485, p=0.746, \text{adj. } R^2=-0.025$				
Variable	B	SE_B	β	p
Intercept	679.05	82.335		<0.001
Component One	-5.26	23.001	-0.026	0.82
Component Two	21.02	22.14	0.105	0.34
Component Three	-6.603	22.31	-0.033	0.76
SRT	15.47	20.78	0.086	0.45

Notes: B = standardised regression coefficient; SE_B = Standard error of the coefficient; β = standardised coefficient. All p -values highlighted in bold are significant at an alpha level of <0.05

Interim Summary

The aims of experiment one were (1) to investigate whether or not participants possess a general ability to adapt to different adverse listening conditions irrespective of the type of distortion; (2) to investigate whether adaptation to the different conditions was dependent on the vocal characteristics of the talker and (3) to uncover the underlying cognitive mechanisms supporting adaptation. Overall the results indicated that participants possess a general ability to adapt in a relatively equal way across all types of distortion, as indicated by the significant correlations between conditions. Secondly, this form of adaptation appears to be dependent on individual differences in vocabulary knowledge, working memory and general cognitive ability (referred to as measures of verbal intelligence). Lastly, both the accuracy and response time results from the generalised linear mixed models indicate a significant effect of speaker, with an overall significantly lower accuracy for speaker four and significantly slower response times for both speaker three and four.

To further investigate the significant effect of speaker, a follow-up experiment was run where only a single speaker was presented per condition, instead of four speakers as in experiment one, i.e., the speakers did not change *within* a condition but did change *between* conditions. This was done to investigate whether the differences in adaptation to the four speakers across conditions found in experiment one would persist if participants heard one specific speaker per condition.

Experiment Two

Methods

All details in experiment two were the same as in experiment one, unless explicitly stated as different.

Participants

Twenty-four different participants (mean age 20yrs 3mths \pm 2.10; range 18-25; 6 males) were recruited. All participants were native British English speakers, had normal or corrected to normal vision, were right handed (Oldfield, 1971; mean score 76.68 ± 19.35 ; range 38.46-100) and were highly educated (mean years of education

15yrs 8mths \pm 1.46; range 14-18). No participants reported a history of speech, language, neurological or psychiatric disorder. All participants gave informed consent and were compensated with monetary payment or course credit.

Procedure

Participants underwent the same pure tone audiometry testing as used in experiment one (octave frequencies between 250 and 8000Hz) in addition to the main speech adaptation task but were not tested on the battery of cognitive assessments. All testing was performed in the same double-walled soundproof room and lasted up to 30 minutes.

Results

Accuracy

Overall accuracy was highest in the clear speech condition (mean = 92.47 ± 26.39), followed by the time-compressed (mean = 81.82 ± 38.58), speech in noise (mean = 63.42 ± 48.19) and finally the noise-vocoded condition (mean = 53.58 ± 49.89).

As in experiment one the results show an overall significant effect of trial number ($\beta = 0.003$, $SE_{\beta} = 0.001$, $p = 0.001$) indicating that performance improved over the trials. However, when analysing the conditions in separate models only the noise-vocoded condition showed an effect of trial number ($\beta = 0.01$, $SE_{\beta} = 0.002$, $p < 0.001$). Furthermore the results show an overall effect of condition, time-compressed ($\beta = -0.09$, $SE_{\beta} = 0.044$, $p = 0.04$), noise-vocoded ($\beta = -0.525$, $SE_{\beta} = 0.05$, $p < 0.001$) and speech in noise ($\beta = -0.383$, $SE_{\beta} = 0.048$, $p < 0.001$) conditions were all performed to a significantly poorer level than the clear speech condition. The effect of speaker was found to be non-significant and there were no significant interactions (see Appendix I for results of follow-up contrasts).

Response Times

Overall participants were quickest to respond in the clear speech condition (mean = 476 ± 399), followed by speech in noise (mean = 713 ± 411), then noise-vocoded (797 ± 507) and finally the slowest overall RTs were in response to the time-compressed speech (871 ± 452). Figure 4 shows the change in accuracy and response times over the course of each condition.

The results show an overall significant effect of trial number ($\beta = -4.867$, $SE_{\beta} = 1.238$, $p < 0.001$) indicating that participants became quicker to respond (i.e., response times went down) as the number of trials increased. When analysing the conditions in separate models, response times in clear ($\beta = -5.16$, $SE_{\beta} = 1.55$, $p < 0.001$), time-compressed ($\beta = -3.57$, $SE_{\beta} = 0.98$, $p < 0.001$) and speech in noise ($\beta = -1.66$, $SE_{\beta} = 0.54$, $p < 0.001$) were found to significantly reduce with increasing trial number; noise-vocoded speech only approached significance ($\beta = -2.36$, $SE_{\beta} = 1.28$, $p = 0.06$). These results largely replicate experiment one and show that participants were able to adapt to the stimuli. Furthermore, a significant effect of condition was found with participants taking significantly longer to respond in the time-compressed ($\beta = 389.01$, $SE_{\beta} = 88.507$, $p < 0.001$) and noise-vocoded ($\beta = 245.87$, $SE_{\beta} = 92.65$, $p = 0.008$) conditions relative to the clear speech condition, but no significant difference between clear speech and speech in noise was found. Lastly, a significant effect of speaker was also found with response times to speaker two ($\beta = 181.52$, $SE_{\beta} = 89.64$, $p = 0.04$) and speaker four ($\beta = 216.24$, $SE_{\beta} = 90.009$, $p = 0.01$) being significantly slower than speaker one.

The two-way interaction between trial number and condition was also found to be significant, as the number of trials increased responses to speech in noise did not improve as much as in comparison to clear speech ($\beta = 4.815$, $SE_{\beta} = 1.4$, $p < 0.001$), there was no significant effect found for the other two conditions across trial number. Finally, the interaction between condition and speaker was also significant with quicker response times for speaker two relative to speaker one in the time-compressed condition ($\beta = -339.19$, $SE_{\beta} = 116.19$, $p = 0.003$; see Appendix I for results of follow-up contrasts).

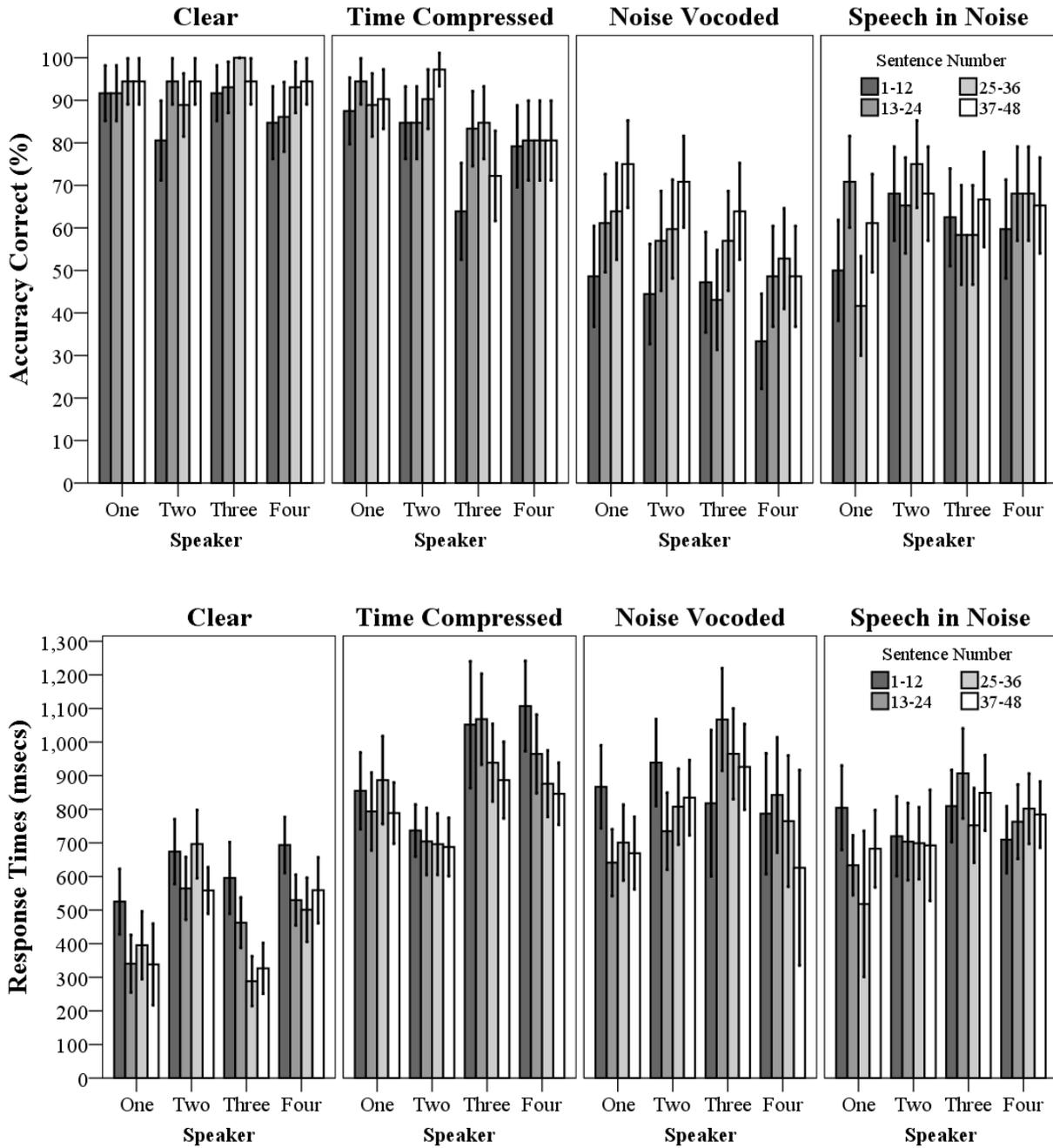


Figure 4 - Accuracy and Response Times for Experiment Two across Condition, Speaker and Sentence Number.

Benefit of Multiple Speakers per Condition: A Comparison of Results from Experiment One and Two

Research investigating perceptual adaptation to foreign accents suggests that high-variability training promotes perceptual learning of distorted speech therefore a comparison was made between the levels of performance per condition across the two experiments. Experiment one adopted the high variability approach using multiple speakers per condition, whereas experiment two used only one speaker per condition. To account for differences in sample size a Welch’s *t*-Test was conducted for accuracy and response times across each condition.

For the accuracy data a significant difference was found between experiments in both the clear speech (Welch’s $t(27.04) = 2.609, p = 0.015$, Cohen’s $d = 0.6$, CI [0.9, 7.54]) and speech in noise conditions (Welch’s $t(37.32) = 3.022, p = 0.005$, Cohen’s $d = 0.695$, CI [2.39, 12.12]). In both conditions, performance was significantly greater in experiment one compared to experiment two. No significant difference was found between levels of performance in the time-compressed or noise-vocoded conditions across experiments.

Table 11 - Showing means and standard deviations for accuracy data across each condition and experiment.

Condition	Experiment	Mean	S.D.
Clear	One	95.97	4.25
	Two	91.73	7.61
Time-Compressed	One	86.10	7.35
	Two	83.94	9.38
Noise-Vocoded	One	60.65	12.95
	Two	54.68	14.11
Speech in Noise	One	70.19	10.65
	Two	62.93	10.37

Notes. Significant differences highlighted by bold text

This pattern of results was also found for the response time data where a significant difference was found between response times in the clear (Welch’s $t(37.39) = -2.12, p = 0.04$, Cohen’s $d = -0.488$, CI [-172.46, -4.65]) and speech in noise conditions (Welch’s $t(48.78) = -3.38, p = 0.001$, Cohen’s $d = -0.77$, CI [-199.91, -50.89]). In both conditions, response times were significantly quicker in experiment one compared to experiment two. No significant difference was found between levels

of performance in the time-compressed or noise-vocoded conditions across experiments.

Table 12 - Showing means and standard deviations for response time data across each condition and experiment.

Condition	Experiment	Mean	S.D.
Clear	One	419.01	184.96
	Two	507.57	179.62
Time-Compressed	One	829.28	204.04
	Two	868.68	197.01
Noise-Vocoded	One	768.47	154.50
	Two	842.21	243.96
Speech in Noise	One	621.52	201.42
	Two	746.93	148.07

Notes. Significant differences highlighted by bold text

Discussion

The present study tested if/how participants adapt to multiple types of speech distortion. Of particular interest was the investigation of whether individual participants performed consistently across the different conditions, i.e., does an individual's ability to adapt vary across different conditions or do they possess (or lack) a general skill to adapt relatively equally to all conditions. Based on the results of Bent et al. (2016) and Borrie et al. (2017), it was expected that participants would be more likely to adapt to all distortions in a similar way. However, in the Bent et al. (2016) study, the stimuli only varied in the quality of their spectral characteristics, whereas the research presented in this thesis represents the first experiments to investigate the effect of a spectral, temporal and environmental distortion on participants performance. Furthermore, these experiments aimed to investigate whether the inclusion of multiple speakers would impact on performance. If adaptation is specific to the distortion then the inclusion of multiple speakers would have little effect, however the inclusion of a high variability training environment could aid learning and lead to greater levels of adaptation. Finally, experiment one aimed to investigate if individual differences in adaptation were associated with individual differences in a battery of audiological and cognitive measures.

Transfer of Learning from one Condition to Another and the Associated Cognitive Mechanisms

One of the most significant findings of the current research are the correlations between performances in all of the conditions. This shows that participants who performed to a high level in one condition, performed to a high level in all speech conditions. This supports the notion that participants possess (or lack) a general ability to adapt to multiple forms of speech, even when the distortions differ in the degree of spectral, temporal or environmental manipulation. With the global skill indicative of a single mechanism (or group of sub-mechanisms) for perceptual learning of degraded speech, i.e., listeners do not invoke a different mechanism or strategy when adapting to spectrally degraded speech compared to when they adapt to temporally degraded speech. This indication of a global skill is in partial support of Bent et al. (2016) who found correlations between most, but not all, of their conditions and full support of Borrie et al. (2017) who found strong correlations between perception of dysarthric speech and speech in noise. Borrie et al. (2017) argue that participants may be employing a metrical segmentation strategy (Cutler & Norris, 1988), this is a strategy whereby the occurrence of a strong syllable, i.e., a syllable containing stress, triggers segmentation of the speech signal whilst the occurrence of a weak syllable does not result in segmentation. Cutler and Norris (1988) propose that using strong syllables to segment the incoming continuous speech signal provides the most efficient location and strategy at which to initiate lexical access. Interestingly, not only did Borrie et al. (2017) find evidence of adherence to the metrical segmentation strategy, with participants making lexical boundary errors which correlated with portions of the speech signal that contained strong stress but were not at the start of a new word. Borrie et al. (2017) also found less evidence for the use of metrical stress as a potential guide to segmentation in the dysarthric speech condition where, as a result of the dysarthria, speech is characterised as having reduced stress. This suggests that when available, one cue that participants might use to segment the distorted speech signal and establish probable locations of word boundaries to assist comprehension is lexical stress. Given the methodology of the current experiment, where participants were simply asked to verify whether each statement was true or false compared to Borrie et al. (2017) where participants were required to write down everything that they thought they had heard, it is not possible to assess whether participants used a metrical segmentation strategy

across the four conditions, however when investigating statistical learning, Palmer and Mattys (2016) suggest that this form of learning was supported by active processes in working memory. Statistical learning refers to the use of regularities within the speech signal to extract words from the continuous speech stream, traditionally this has been investigated in the context of language acquisition however the same principles would apply during perception of degraded speech as well. In addition to working memory, higher levels of statistical learning have also been associated with an overall higher level of vocabulary knowledge (Spencer, Kaschak, Jones, & Lonigan, 2015). In the current experiment, individual differences in performance in the clear, time-compressed and speech in noise conditions were all associated with individual differences in performance on tests of working memory, vocabulary knowledge and general cognitive ability (using the MoCA which also tests working memory and vocabulary knowledge). Given the nature of the task and the association of working memory and vocabulary knowledge with statistical learning it is likely that participants were using a form of this learning to assist segmentation and subsequent adaptation to the distorted stimuli. This hypothesis is supported by Neger et al. (2014) who found a significant effect of vocabulary knowledge on perceptual learning of noise-vocoded speech. Additionally, Neger et al. (2014) found that perceptual learning was modified by statistical learning ability, with those participants who showed better performance on a statistical learning task also showing greater perceptual learning of noise-vocoded speech (in their younger group of participants, but not in the older group). The authors argue that perceptual learning abilities may rely directly on sensitivity to the probabilistic information inherent in all speech (i.e., statistical learning). It is believed that individuals who are more capable, and faster, to identify subunits of the distorted speech signal are able to transfer this information to higher level processors thus facilitating faster access to lexical representations and greater overall adaptation to the distortion. Whether or not metrical segmentation was being used as the main cue for statistical learning cannot be answered with the current data. Clear, time-compressed, and speech in noise all retain the suprasegmental cues necessary for metrical segmentation to occur, however Sebastián-Gallés et al. (2000) found that adaptation to time-compressed speech is dependent on the isochrony of the habituating language. Therefore, rather than using stress timing as the predominant cue in adaptation it is more likely that individuals use the specific isochrony of the language that they are exposed to as a predominant cue for statistical learning and adaptation, a process that

appears to be supported by individual differences in working memory, vocabulary knowledge and general cognitive ability.

Perceptual Adaptation to the Different Speech Distortions

Of the four speech conditions included in the two experiments of this chapter, overall performance was highest in the clear speech (as would be expected), followed by time-compressed, then speech in noise and finally noise-vocoded speech. However, the level of adaptation in both experiments was only visible in the response time data where participants became significantly quicker at accurately verifying the content of heard sentences. No significant improvement was observed in the accuracy data in experiment one and only the noise-vocoded condition produced learning effects in experiment two. This is not the first set of experiments to find little to no effect of exposure on accuracy in learning experiments (Adank & Devlin, 2010; Peelle & Wingfield, 2005) and the current results replicated the findings of Bent et al. (2016) who also found no adaptation when participants were presented with a variety of speech distortions in the same experimental session. This is likely to have occurred as a result of different factors; firstly, it is possible that the specific manipulations employed in the current study were either too lenient or too strenuous to allow significant improvements in recognition accuracy. In the current study, the time-compressed speech was shortened to 40 percent of its original length, and whilst this is similar to other studies (Dupoux & Green, 1997; Golomb et al., 2007; Peelle & Wingfield, 2005; Sebastián-Gallés et al., 2000) all of which compressed speech to between 30 and 45 percent of its original length, previous studies all used word report scores as the main dependent variable. However, the experiments presented here used a sentence verification task which restricts minimum performance to 50 percent (chance level, unless participants actively attempt to answer trials incorrectly in which case, performance can fall below 50 percent). It is possible therefore that a combination of a compression rate of 40 percent and use of a task with a high level of chance performance resulted in participant performance peaking from the very start and therefore unable to improve further. In contrast, the noise-vocoded speech condition was vocoded into four logarithmically spaced channels with a cut-off frequency of 5000Hz. This is more stringent than previous studies, most of which use

six channels to vocode their speech stimuli (Davis et al., 2005; Hervais-Adelman et al., 2012; Hervais-Adelman et al., 2008; Huyck & Johnsrude, 2012). It is possible therefore that, unlike the time-compressed condition where the manipulation employed could have been too lenient given the different task demands, the four channels used to create the noise-vocoded speech in the current set of experiments could have been too difficult for participants to adapt to. Indeed Loizou et al. (1999) noted that adaptation to noise-vocoded speech drops rapidly below five channels. Although adaptation in terms of improvements in accuracy were observed for the noise-vocoded speech condition in experiment two, which impacts on the strength of this argument as it suggests that participants were capable of adapting to noise-vocoded speech despite the difficult parameter settings.

Alternatively, D. R. Moore, Amitay, and Hawkey (2003) distinguish between procedural adaptation and perceptual adaptation. Procedural adaptation is any change in performance related to participants becoming used to the task demands and developing a full understanding of the rules of the task. Perceptual learning however is independent of the task and results in a change in the ability of the participant to detect, discriminate or identify an acoustic stimulus after a period of exposure despite initial attempts to do so being unsuccessful (Watson, 1980). It is possible that procedural learning is more associated with accuracy of performance whilst perceptual learning is more closely linked to changes in response times. In the current experiments, all participants were presented with the clear speech condition first specifically to ensure that all adaptation related to procedural learning effects were removed from the results of the three speech distortion conditions. It is possible that in doing so, participants became overly familiarised with the task and setting before any speech distortions were heard. This could have resulted in a higher starting accuracy for the speech distortions which in combination with the too lenient/stringent distortions could have also limited the amount of improvement that was possible in each condition. Future studies will be needed to investigate if the removal of a clear speech baseline or moving the baseline to different positions in the order impacts on changes in recognition accuracy.

Additionally, feedback was not provided to participants at any point during the two experiments presented here, however this has been found to induce greater levels of adaptation for speech in noise (Burk et al., 2006) and noise-vocoded speech (Davis

& Johnsrude, 2003; Hervais-Adelman et al., 2008; Loebach et al., 2010). These conditions produced overall the lowest accuracy levels across the two experiments and it is possible that this could be increased with the provision of written or aural feedback of either the sentence in clear form or simply a statement of accuracy of performance (i.e., Correct or Incorrect).

Finally, as noted above, significant reductions in response times were observed across all four speech conditions in both experiments one and two, a result that was interpreted as evidence of *perceptual* adaptation to the different speech distortions. However, no significant speech condition by trial interaction was observed in either experiment, which suggests that the rate of adaptation across all four conditions was both quantitatively and qualitatively equivalent. A qualitative equivalence in adaptation rate between the conditions would either suggest that participants also had to perceptually adapt to the clear speech condition, which is unlikely given the ease of perceiving stimuli in this condition, or that the adaptation observed in the three distorted conditions does not represent genuine perceptual adaptation but is instead more reflective of an ongoing task-related practice effect. Whilst it is hard to entirely refute such a claim given the observed data, it is hypothesised that the nature of adaptation between conditions is quantitatively but not qualitatively equal. Instead it is believed that the reduction in response times in the clear speech condition represents what Moore et al. (2003) would term procedural learning i.e. adaptation to the task demands, whilst the improvement in the time-compressed, noise-vocoded and speech in noise conditions is qualitatively different and is more representative of perceptual learning i.e. due to changes in weighting at either the phonological (for time-compressed) or semantic level (for noise-vocoded and speech in noise) and not due to changes in task familiarity. This hypothesis is based on the fact that in both experiments the clear speech condition was always the first condition that participants were exposed to, with the methodological reasoning that this would provide the participants with an opportunity to adapt to and overcome any task-related practice effects (i.e. procedural learning) before exposure to the distorted stimuli. If this manipulation was successful, then the participants should have fully adapted to the task by the end of the clear speech condition with all subsequent improvements in performance being dependent on perceptual adaptation to the specific distortions rather than further procedural adaptation. Future research could investigate this

hypothesis by increasing the number of trials per condition to investigate whether the hypothesised qualitatively different adaptation profiles would emerge with further testing. Alternatively, a fifth condition could be introduced using a type of distortion that is not expected to result in any form of adaptation within the short time-window of this experiment, for example, spectrally-rotated speech which has been shown to require many hours of training before adaptation occurs (Green, Rosen, Faulkner, & Paterson, 2013) or a second block of clear speech. If, as hypothesised, all procedural learning occurs before the end of the first block of clear speech and all subsequent adaptation in the time-compressed, noise-vocoded and speech in noise conditions is dependent on perceptual adaptation, then no significant reduction in response times would be expected during exposure to either the spectrally-rotated speech (which is believed to require exposure to more than 48 sentences before perceptual adaptation is possible) or during the second presentation of clear speech (which should be easy enough to perceive for native speakers that no perceptual adaptation would be required). If such a result was observed in combination with a replication of the significant reduction in response times for the clear speech, time-compressed, noise-vocoded and speech in noise conditions from experiments one and two, then it would support the hypothesis that the underlying cause of the improved performance in the clear speech condition i.e. procedural learning, was qualitatively different to the underlying cause of the improved performance observed in the time-compressed, noise-vocoded and speech in noise conditions, i.e. perceptual learning. If, however, the spectrally-rotated and/or second clear speech condition also resulted in a significant reduction in response times, in addition to a replication of the findings from experiments one and two, then it would suggest that the underlying cause of the improved performance in the first two experiments of this thesis was qualitatively equivalent across conditions and therefore most likely not representative of genuine perceptual adaptation.

Transfer of Learning from one Speaker to Another

The inclusion of multiple speakers for all conditions in experiment one produced two notable results. If perceptual learning of a speech distortion was independent of speaker then no significant differences would be expected between speakers.

However, speaker four on the whole was found to be less intelligible, in both accuracy and response times measures across both experiments whilst partial difficulties in adapting to speaker three (response times in experiment one) and speaker two (response times in experiment two) were also found. This supports the notion that adaptation was dependent not only on the condition but also on the speaker. This has important ramifications firstly for the field as a whole where the majority of studies use a single speaker during adaptation. It would be impossible to establish whether the presence or absence of an adaptation effect was due to the characteristics of the distortion, the speaker or both. More importantly, a result which shows that adaptation is speaker as well as distortion dependent, has ramifications in everyday life. On a lower level, if learning interacts with vocal characteristics of the speaker then this might explain, for example, why some people with a foreign accent remain hard to understand, while others are easy to adapt to. Yet, on a more important clinical level, noise-vocoded speech is believed to simulate the experience of using a cochlear implant. If adaptation is dependent on vocal characteristics as well as distortion, then this would suggest that there are certain individuals for whom cochlear implant users will find explicitly harder to perceive and adapt to. As the results also suggest that individuals possess a global skill for adaptation then successful adaptation to use of a cochlear implant would be dependent not just upon the technological settings of the implant but also upon the vocal characteristics of encountered speakers and the degree to which each individual user possessed or lacked the “global skill” to adapt. Research thus far has suggested that cochlear implant users are unable to accurately discriminate between different speakers, with accuracy performance at 23 percent correct (although chance level performance equated to 10 percent, i.e., 10 speakers), equating to a level of performance that 1 channel noise-vocoded speech would produce in healthy subjects (Vongphoe & Zeng, 2005). In the current experiment, participants were not asked to identify or discriminate between speakers and were not informed that sentences from multiple speakers would be presented (although presentation of the clear sentences first would have made this experimental manipulation obvious to participants once the experiment began) and therefore inferences on whether or not participants actually distinguished between participants must be tempered. However, whether or not participants would be able to identify the individual speakers does not detract from the finding that overall perception of speaker four was significantly less than for the other speakers when the speakers were intermixed (experiment one) or

held constant (experiment two). Therefore, whether participants were consciously tracking individual speaker characteristics or not, most likely not in experiment two where only a single speaker was presented per condition, there appear to be certain characteristics of each talker which potentially subconsciously appear to aid or impair adaptation to that speaker's voice. The results of the current experiments suggest speaker one spoke at an overall higher pitch and slower speaking rate than the other three speakers both of which could be influential in determining why participants generally performed well when sentences were spoken by this speaker. Indeed, Palmer and Mattys (2016) found that segmentation was improved when the speech rate of their stream of nonsense words was reduced. The authors argue that as statistical learning is dependent on working memory processes, slowing down the speech rate places less strain on working memory capacities allowing greater analysis of the speech stream. The slower overall syllable rate of speaker one may therefore be one of the factors contributing to overall high levels of adaptation for this speaker. Future research will be needed to establish the importance of speaking rate and what other specific speaker characteristics are influential in causing differential rates of adaptation between speakers.

The second effect of using multiple speakers was that it appears to lead to a significantly higher level of performance for speech perceived in clear and noise backgrounds (as shown by Welch's *t*-Test of performance in experiment one to two) than when only a single speaker is perceived. In addition, although not significantly so, better performance was also observed for time-compressed and noise-vocoded speech when multiple speakers were heard during adaptation. This result replicates research in training of foreign/non-native phonemic contrasts which repeatedly show that high variability training leads to a greater level of adaptation and acquisition (Bradlow & Bent, 2008; Shinohara & Iverson, 2018; Wang et al., 1999). Shinohara and Iverson (2018) argue that training on highly variable stimuli does not lead to a remapping of the cues used to identify speech sounds, in an attempt to make them more appropriate for the distortion being heard but instead the training makes participants more consistent and automatic in their application of cues they possessed before training. In addition, Shinohara and Iverson (2018) suggest that whilst this technique leads to some improvement in identification/discrimination, without any correction to the mistuned phonetic processing, improvements made during the high-variability may

peak after a small amount of improvement. The current study reports greater performances for high-variability training for the speech in noise and clear speech conditions. It is interesting that these are the two conditions that most closely resemble real world situations e.g. when following the conversation of a group of friends in a busy café or listening to a discussion amongst colleagues in a quiet meeting room. It is likely therefore that these are conditions in which participants had vast amounts of experience of perceiving and switching between multiple speakers and therefore had more fine-tuned and appropriate strategies to apply and aid perception. In contrast, attempting to use the same acoustic cues to perceive speech that has been distorted in a non-familiar way, i.e., time-compressed or noise-vocoded leads to less improvement.

Conclusions

In conclusion, results from the current set of experiments indicate that individual listeners possess a general skill to adapt to various speech distortions. This skill allows participants to adapt to a relatively equivalent level across different conditions irrespective of whether the distortion is temporal, spectral or environmental in nature. Equivalent adaptation across distortions is indicative of the use of a similar (and possibly singular) cognitive-perceptual processing mechanism in all adverse listening conditions. The results of this experiment suggest that measures of verbal intelligence, specifically vocabulary knowledge and working memory as well as general cognitive functioning could underpin the perceptual learning process and provide support for the requisite statistical learning to occur to assist adaptation. Furthermore, whilst the perception of and adaptation to certain speakers was found to be less successful than to other speakers, use of high variability training produced an overall greater level of performance in the speech in noise and clear conditions (with numerical improvements also in time-compressed and noise-vocoded). This has important implications for future research in two ways, firstly, future research could seek to establish what specific acoustic characteristics of the speaker either aid or impair adaptation. Secondly, future research could establish whether the benefit of high-variability training extends to the more extreme forms of distortions, such as time compression or noise-vocoded, or whether such benefits are limited to the more real world, highly experienced clear, accented and speech in noise conditions. Overall, the current

research adds to the work of Bent et al. (2016) and Borrie et al. (2017) in providing a more comprehensive overview of the adaptation characteristics of individual listeners to multiple distortions and multiple speakers, an area of the field that is currently under researched and therefore lacking in knowledge.

Chapter Three

Investigating the Efficacy of TMS as a Research tool to Disrupt Auditory Sentential Processing (in Noise).

Introduction

In 1874, after observing patients who exhibited rapid and effortless speech production with somewhat intact hearing but a profound loss of ability to comprehend speech, Carl Wernicke stated that the posterior superior temporal gyrus (pSTG) must be the sensory speech centre of the brain. In the intervening 144 years, the notion that the pSTG is the neural locus of auditory speech processing has become one of the most widely accepted concepts in cognitive neuroscience. However, fundamental questions relating to the neurobiology of speech perception still exist. In the following section, I will provide a brief overview of the debate as it stands, before discussing how TMS could be used effectively to address some of the remaining questions.

At present the two most prominent neurobiological models of speech perception are the principally unilateral model of Rauschecker and Scott (2009) and the bilateral model of Hickok and Poeppel (2000). Both of these models build on the notion that speech perception occurs in the context of a dual stream of processing with a ventral pathway involved in mapping sound to meaning and a dorsal pathway mapping sound to articulatory motor processes. However, based mainly on nonhuman primate neuroanatomical data, the Rauschecker and Scott (2009) model argues that “...*speech perception and production are left lateralised in the human brain.*” (p. 720) with the locus of successful speech perception in the left anterior superior temporal sulcus (STS) (Rosen et al., 2011; Scott et al., 2000). In contrast, the Hickok and Poeppel (2000) model (based mainly on data from clinical human populations) argues that speech perception occurs bilaterally. Earlier versions of this model argue that bilateral processing only occurs at the pre-lexical level with all subsequent lexical and post-lexical processing being processed in the left hemisphere. However, more recent versions posit the existence of a pathway in each hemisphere, which is capable of processing speech sounds up to and including the mental lexicon (Hickok & Poeppel, 2007). In this model, successful speech perception relies on sites both anterior and posterior to the transverse temporal gyrus with phonological processing especially

occurring bilaterally and semantic processing being more left dominant (Hickok, 2009).

The argument in favour of bilateral speech perception comes mainly from patient data where unilateral lesions or unilateral anaesthetisation during Wada testing (Wada, 1949) of either the left *or* right hemisphere (Hickok et al., 2008; McGlone, 1984) partially impairs speech perception suggesting both hemispheres are important for this task. Additionally, patients with bilateral lesions encompassing both left *and* right superior temporal regions are more likely to suffer from verbal auditory agnosia, a condition defined as having an inability to understand spoken language despite preservation of other language capabilities, i.e., reading or writing (Buchman, Garron, Trost-Cardamone, Wichter, & Schwartz, 1986). Of 63 well-detailed cases of verbal auditory agnosia roughly 70 percent involved bilateral lesions in agreement with Hickok and Poeppel (2000), and supporting the notion that both left and right hemispheres are critical to speech perception. Yet the remaining 30 percent had unilateral damage and only one patient had a right hemisphere lesion (Slevc & Shell, 2015). Additionally, not all studies find a negative effect on speech perceptual abilities following right hemisphere anaesthetisation during Wada testing (Boatman et al., 2000; Boatman et al., 1998). This suggests that whilst the right hemisphere is important for speech perception it does not play a critical role, unlike the left hemisphere. Results from patient data are therefore inconclusive with respect to the unilateral or bilateral organisation of speech perception. A reason for this is that the damage caused by lesions or strokes is not constrained to functional or anatomical boundaries and following injury neural plasticity is important in brain reorganisation and repair. This inconsistency in lesion location or reorganisation across patients makes it difficult to conclusively localise the cortical origin of specific cognitive functions and thus it makes it harder to predict the resultant cognitive deficit.

With the advent of functional imaging came the ability to examine the processing of the healthy, intact human brain during completion of the relevant task. However, neuroimaging studies on healthy participants have also reported mixed results with respect to the neurobiology of speech perception. Early neuroimaging studies on processing of intelligible speech reported a left-lateralised anterior temporal lobe locus of processing (Narain et al., 2003; Scott et al., 2000), while later studies have reported bilateral involvement of anterior and posterior temporal areas in speech

perception (Evans, Kyong, Rosen, Golestani, Warren, McGettigan, Mourão-Miranda, et al., 2014; Friederici et al., 2010; Harris et al., 2009; Obleser, Eisner, & Kotz, 2008; Okada et al., 2010; Rosen et al., 2011; Zekveld et al., 2006). It is likely that the lack of a coherent conclusion from these studies is due to both methodological differences and the correlational nature of such methods where observed changes in the BOLD signal can either be functionally relevant or epiphenomenal.

All methodological techniques contain inherent limitations that impact upon the potential conclusions that can be drawn. In order to gain the most comprehensive and accurate insight into the cortical structures and processes underlying particular cognitive functions, it is advisable that multiple research techniques should be used to investigate the same process. Through the use of multiple methodologies, the knowledge gained from the strengths of one technique can negate the gaps left by the weaknesses of others (Bechtel, 2002). The use of patient data provides an insight into the effects of damage to the brain on cognitive functioning, allowing inferences related to the role that the damaged regions play in behavioural tasks. Yet, this data is biased by neural plasticity and large between patient variation in the location and extent of damage. Whilst functional imaging provides an insight into the *in vivo* workings of the human brain during completion of the task of interest, however research using this technique is limited by its correlational nature.

Transcranial Magnetic Stimulation (TMS) is a neurophysiologic technique that allows for non-invasive stimulation of the human brain through the application of strong but short electrical pulses that enable us to modulate the underlying neural activity in conscious, healthy human subjects (non-invasively). By inducing electrical currents in the brain that modulate and disrupt the ongoing activation within a given region, TMS can be used to demonstrate causality between a cognitive process and specific brain regions and as a result can be used to complement other neuropsychological techniques (such as fMRI, EEG), which are purely correlation in nature. In TMS, the pulse is sent along the TMS coil, reaching its peak and returning to zero in less than a millisecond, the very rapid nature of the pulse induces a magnetic field perpendicular to the plane of the coil that also rises/falls rapidly in time. The rapidly fluctuating magnetic field passes unimpeded through the scalp and skull of the participant and induces an electrical current in the brain. If the induced current is of sufficient intensity, it will depolarise the neurons in the targeted region, therefore the

overall effect of a TMS pulse (or a train of pulses) is one of cortical excitation. However, depending on the region and the type of stimulation (i.e., the timing, duration and frequency) this cortical excitation can have either an inhibitory or excitatory effect on task performance (Adank, Nuttall, & Kennedy-Higgins, 2017; Fitzgerald, Fountain, & Daskalakis, 2006). The pulses can occur in isolation, as a pair, or in trains, with administration of more than a single pulse referred to as repetitive TMS (rTMS). Additionally, pulses either occur online, at the same time that participants are performing the task or offline, with pulses being administered before the task begins (Adank et al., 2017). The focality of a single pulse of TMS is usually measured on the cortical surface and depends on the type of coil but is estimated to have a surface spatial resolution of 5-20mm (Deng et al., 2013). Such spatial precision is advantageous in comparison to patient data where the effects of brain trauma are often more diffuse and vary greatly across patients thus allowing more accurate interpretations of the functional relevance of specific brain regions on a task of interest (Sliwiska, Vitello, & Devlin, 2014). However, when the cortex is stimulated, there is not only a change in activation in the targeted region, but also a change in activity in immediately surrounding areas of cortex as well as in more distally connected cortical areas. Therefore, whilst TMS is initially spatially precise, it does afford the opportunity to investigate the impact of a specific modulation on larger cortical networks. Given the potential for causal conclusions, TMS is theoretically the perfect methodological addition to the neurobiology of speech perception field. It is therefore surprising that very few related studies have been published investigating this topic (see Table 13 for an overview of studies applying TMS to the temporal lobes during a speech perception task).

Thus far TMS has been found to impair both semantic and phonological judgments after left posterior STG (pSTG) stimulation (Krieger-Redwood, Gaskell, Lindsay, & Jefferies, 2013); as well as prosodic judgment (Alba-Ferrara, Ellison, & Mitchell, 2012) and human voice perception after right pSTG stimulation (Bestelmeyer, Belin, & Grosbras, 2011). Given the critical importance of these regions in speech perception it is perhaps not surprising that the application of TMS disrupted performance across these studies. Krieger-Redwood et al. (2013) conclude that the impairment is the result of TMS increasing the “ambiguity of the auditory input to the system which necessarily impacts on processing at all levels” (p.2185). Taken together

this research suggests that TMS can be used to further our understanding in a way that is complementary to other research techniques.

Yet, TMS also has limitations that need to be considered before adopting it as a viable technique. The most referred to concept in TMS research is the creation of a ‘virtual lesion’ (Pascual-Leone, Bartres-Faz, & Keenan, 1999). A common misunderstanding of this phrase is that the induced ‘virtual lesion’ results in a complete loss of cognitive ability within the region being stimulated, i.e., TMS is capable of inducing deficits akin to cortical deafness. Whilst some such effects have been observed in the visual system (Amassian et al., 1989) generally the effects in other cortical regions are far subtler and experiments rely on more fine-grained distinctions in performance across tasks/stimulation sites on the order of millisecond and microvolt changes in responses. This is a general limitation of TMS rather than a specific obstacle to neurobiology of language research, however it must be taken into account when assessing the research conducted so far. Meister, Wilson, Deblieck, Wu, and Iacoboni (2007) found no effect on discrimination of two consonant-vowel syllables in noise after left STG stimulation despite finding an impairment of tone discrimination and Drager, Breitenstein, Helmke, Kamping, and Knecht (2004) found no effect relative to baseline in a picture-word verification task. Whilst Beauchamp, Nath, and Pasalar (2010) found that subjects were significantly less likely to report the McGurk effect after single pulse TMS of the STS. However, they conclude that this result is best explained as interfering with audio-visual integration rather than as evidence that TMS can interfere with speech perception. The differing results outlined here compared to those discussed previously highlight that whilst TMS does provide the opportunity to establish causal brain-behaviour links the very subtle effect that TMS has on the overall network makes the task far more complex. Indeed, Meister et al. (2007) theorise that the network for speech perception within the temporal lobes is too extensive to be compromised by TMS “because of compensatory processes within the contralateral temporal cortex” (p.1695).

This view is supported by Andoh and Paus (2011) who combined 1Hz offline rTMS with functional imaging to investigate the impact that stimulating the posterior superior temporal region of each hemisphere would have on activation in the contralateral hemisphere. The results showed a task related increase in activation in the homologue areas contralateral to the site of stimulation, i.e., stimulation of the left

posterior temporal region resulted in a task related increase in activation in the right STG/MTG and the left cerebellum. Andoh and Paus (2011) suggest that these results are evidence of the brain compensating (within four minutes of stimulation) for the TMS induced disruption to one hemisphere by drawing in additional resources from the opposite hemisphere. The authors suggest that this interhemispheric compensatory process is the reason why behavioural effects are not always observed after application of TMS (similar neurological adaptation effects were also observed after application of rTMS on a reading task, Mason, Prat, and Just (2014)) and further suggest that the interhemispheric compensation observed is likely to represent the early stages of the longer term and permanent processes that occur in patients following neurological trauma with individual differences in the degree of interhemispheric compensation explaining the variable impact of unilateral or bilateral damage.

The following series of experiments therefore represent an attempt to find the most effective TMS protocol to non-invasively impair sentential speech perception in noise in healthy human adults. In experiments three to six, participants are asked to perform a series of speech reception threshold (SRT, Plomp & Mimpen, 1979a, 1979b) tasks whilst receiving TMS to one of three cortical locations: left STS, right STS and a control site (vertex is used as a control site in experiments three and six; occipital pole in experiment four and the lateral occipital complex in experiment five) as well as completing an SRT task in a baseline no TMS condition. The SRT task assesses participants' ability to perceive sentences in noise. Participants were asked to report up to five key words, and the background noise level was varied dependent on their performance. If they could name three or more key words, the background noise level increased, making the task harder. If they named fewer than three key words, the noise level was decreased in amplitude relative to the sentences, thus making the task easier. The dependent variable of the SRT is the average signal-to-noise ratio at which participants are able to perceive 50 percent of the key words. If the targeted area is causally involved in speech perception, and if the TMS protocol being used is effective, then rTMS to this region should result in an increased SNR, indicating that participant could tolerate less noise in order to perform the task. In experiments three and four a Gap Detection Threshold (GDT) task was conducted, mainly as a control task, in conjunction with the SRT task to investigate the level of auditory processing, if any, that was disrupted by the application of TMS. On the GDT task, participants

heard three separate bursts of white noise. In the middle of one of the bursts there is a gap (a period containing no sound) of varying length. Participants were asked to indicate which of the bursts they believed contained the gap, as with the SRT, the length of the gap varies depending on performance on the preceding trial. Finally, in experiment seven participants were asked to complete a visual discrimination threshold task under the same four TMS conditions as used in experiment six (left STS; right STS; vertex; no TMS). A visual discrimination threshold task was used, as it is a task for which the bilateral STS is functionally irrelevant, i.e., the bilateral STS is not involved in silent reading of letter strings. Use of this task was intended to investigate the validity of the results obtained in experiment six. All experiments used repetitive pulses, experiments three to five adopt an offline rTMS protocol whilst experiments six and seven utilised an online rTMS protocol.

Table 13 - Summary of previous research studies investigating the effect of TMS to temporal cortical regions.

Authors	TMS paradigm	N	Online/Offline	Site stimulated	TMS Intensity	Dependent variable	Task	Site locator	Results
Alba-Ferrara et al. (2012)	1Hz; 60% of max. output and held constant; 10 minutes; 600 pulses	11	Offline	pSTG (bilaterally)	60	Accuracy; Reaction times	Prosody decoding; Semantic judgment	Anatomically	Accuracy – no effect; Reaction times –Prosody trials negatively affected after right pSTG; Incongruent semantic trials bilaterally affected compared to baseline
Andoh et al. (2006)	1Hz; 110% of rMT; 10 minutes; 600 pulses	11*	Online	Left pSTG; Left Pars Opercularis	65	Reaction times	Fragment detection task	MRI functionally located	pSTG – facilitation; Pars Opercularis – no effect
Andoh et al. (2008)	Same as above preceded by either: 1Hz offline; 110% of MT;	13*	Both	Left pSTG	63	Reaction times	Fragment detection task	MRI functionally located	1Hz priming – native language facilitation; iTBS priming – native and

Authors	TMS paradigm	N	Online/Offline	Site stimulated	TMS Intensity	Dependent variable	Task	Site locator	Results
	10 minutes; 600 pulses; Or iTBS; 90% of MT; 3 pulses at 50Hz repeated with 200ms gaps for 2 seconds repeated every 10 seconds; 600 pulses.								non-native facilitation
Andoh and Paus (2011)	3 times (15x 10Hz trains with ITI of 10 seconds); IBI of 10 mins; 450 pulses	20	Offline	pSTG (bilaterally)	63, 66	Reaction Times; BOLD signal	Fragment detection task	MRI functionally located	Reaction times – facilitation after left pSTG BOLD – increase in contralateral area post TMS
Beauchamp et al. (2010)	Single pulse; 100% MT	12*	Online	Left pSTS	68	Accuracy	McGurk effect	MRI functionally located	Left pSTS – suppression of McGurk effect. No effect on simple speech

Authors	TMS paradigm	N	Online/Offline	Site stimulated	TMS Intensity	Dependent variable	Task	Site locator	Results
Bestelmeyer et al. (2011)	10Hz; 110% of rMT; 4 pulses	9	Online	rTemporal Voice Area; rSMG	58	Accuracy	Discrimination	MRI functionally located	perception (control task) rTVA – disruption of voice/non- voice task, no effect on intensity
Drager et al. (2004)	1Hz; 110% of rMT; 10 minutes; 600 pulses	20	Offline	F7 (IFG), F8, Cp5 (pSTG), Cp6, Oz, sham	60	Reaction times	Picture- auditory word verification	International 10-20 electrode system	Cp5 – no significant effect F7, F8, Cp6, sham – facilitation
Grabski, Tremblay, Gracco, Girin, and Sato (2013)	Single pulse; 110% rMT	12	Online	Left: pSTG SMG PMv	70	Accuracy; Reaction Times	Syllable Identification / Categorisation	Anatomically	Reaction times – slowed for pSTG and SMG relative to sham stimulation. Accuracy – equal adaptation effect across all three sites.

Authors	TMS paradigm	N	Online/Offline	Site stimulated	TMS Intensity	Dependent variable	Task	Site locator	Results
Hirnstain, Westerhausen, and Hugdahl (2013)	1Hz; 110% of MT; 10 minutes; 600 pulses	18	Offline	Planum Temporale (bilaterally)	55, 57	Accuracy; Laterality quotient	Dichotic listening task	Not explicitly stated	Accuracy – No effect; Laterality Quotient – rPT facilitation
Krieger-Redwood et al. (2013)	1Hz; 120% of aMT; 10 minutes; 600 pulses	15*	Offline	Left pSTG; Left PMC; Occipital Pole (control)	49	Reaction times	Discrimination	Co-ordinates	pSTG – Suppression of both phonological and semantic task; PMC – suppression of phonological task OP – no effect
Meister et al. (2007)	1Hz; 90% of rMT; 15 minutes; 900 pulses	19	Offline	Left pSTG; Left PMC		Accuracy	Discrimination	MRI functionally located	pSTG – no effect on speech task, suppression only on tone task PMC – only a suppression on speech task

Notes: * represents cases where participants were removed from the initial dataset prior to analysis. Initial number of recruited subjects will be higher, values reported are the number of participants included in the final analysis.

Experiment Three

Methods

Participants

Sixteen participants (mean age 23 years \pm 2.87; range 19-29; 12 females) were recruited for this experiment. All participants were native British English speakers, had normal or corrected to normal vision and were right-handed as assessed through dominant writing hand. No participants reported a history of speech, language, neurological or psychiatric disorder. All participants were assessed to have normal hearing, i.e., average pure tone threshold of 20dB HL or better at octave frequencies between 250 and 8000Hz in both ears (British Society of Audiology, 2011). No participants presented with any contraindications for either MRI or TMS, all participants gave informed consent and were paid for their participation.

Procedure

To evaluate the efficacy of TMS as a research tool capable of disrupting higher level auditory processing of sentential speech stimuli presented in noise, participants were asked to complete four separate SRT tasks (Plomp & Mimpen, 1979a, 1979b) and four gap detection threshold tasks. In three of the four tests, participants received TMS to either the left or right superior temporal sulcus or the vertex. In the other test, no TMS was administered in order to establish baseline performance, order of stimulation site was counterbalanced.

Speech Reception Threshold

All sentences occurred in the presence of speech shaped noise with the SNR varying adaptively depending on individual participant performance. The first sentence was presented at an SNR of +20dB. Correct repetition of three or more keywords resulted in a reduction of 10dB on subsequent trials, until participants were unable to correctly repeat more than two keywords. At this point the SNR increased in steps of 6dB until another reversal occurred (i.e., participants' correct repetition of three or more of the key words) with all subsequent changes occurring in steps of 4dB. A reversal refers to

the shift in direction of SNR change from one trial to the next, for example, if a participant repeated more than three key words for four sentences in a row then the SNR will reduce after each sentence making the subsequent sentence on each occasion harder to perceive. If on the fifth sentence the participant was unable to repeat at least three of the key words, the SNR will increase making the subsequent sixth trial easier to understand. This change in direction from decreasing to increasing (or vice versa) SNR represents a ‘reversal’. Participants’ speech reception thresholds (SRT) were computed by taking the mean SNR (dB) from all trials where a reversal occurred (Plomp & Mimpen, 1979a, 1979b).

After presentation of each sentence, participants were asked to repeat verbatim what they heard. Responses were scored online immediately after each trial using a graphical user interface (GUI) on a standard computer screen that was not visible to participants. Within each sentence there were five key words upon which scoring was based, for example ‘The MEAL was COOKED BEFORE the BELL RANG’ (keywords in uppercase letters). In addition to perfect repetition, keywords were adjudged to be correct if participants changed the grammatical number of presented words e.g. ‘Meals’ (plural) instead of ‘Meal’ (singular). All other digressions were scored as incorrect with no feedback given to the participants.

Orders of sentence list were counterbalanced using a Latin-square technique. All sentences were pseudo-randomly ordered such that the order of presentation was different between participants but each sentence was only played once per participant.

Gap Detection Threshold

On each of the GDT trials, participants heard three separate bursts of white noise. In the middle of one of the bursts there was a gap (a period containing no sound) of varying length. Participants were asked to indicate which of the bursts they believed contained the gap by selecting one of the three corresponding visually presented stimuli (using a standard P.C. mouse). The bursts of white noise were separated from each other by 200 milliseconds of silence. On the first trial of the GDT task the gap is 30 milliseconds in length and therefore is relatively easy to perceive. The length of the gap varies adaptively thereafter depending on the accuracy of the participant in detecting the gap (Levitt, 1971). There were a maximum of 30 trials per block, the

dependent variable for the GDT test is the average length in milliseconds of the gap for the final four reversals in a test. Half of the participants were randomly allocated to complete the GDT tests before the SRT test, whilst the other half performed the SRT test before the GDT test.

Stimuli

Speech Reception Threshold

Four lists of 30 sentences were created from a pre-recorded set of IEEE sentences (see Appendix J). The same male speaker of standard southern British English read all sentences in a sound attenuated room. Audio digitising was performed at 44.1kHz and 16 bits. The beginning and end of each sentence was trimmed to zero crossings as closely as possible to the onset/offset of the initial and final speech sounds. The sentences were then peak-normalised to 99 percent of maximum amplitude and scaled to 70dB SPL using Praat (Boersma & Weenink, 2011). Sentences were presented in steady-state speech-shaped noise, the spectrum of which was derived from the 120 test sentences. On all trials, the noise masker started 500 milliseconds before the onset of the sentence. All sentences were presented binaurally via Etymotic ER1 earphones using a custom-made MATLAB 2014a script (The MathWorks Inc., Natick, MA, 2000).

Gap detection threshold

41 bursts of white noise were created containing gaps that varied logarithmically in length from stimulus one (gap length = 0.5 milliseconds) to stimulus 41 (gap length = 30 milliseconds). All stimuli were 300 milliseconds in total duration, irrespective of the length of gap. All stimuli were presented binaurally using the same set of Etymotic ER1 earphones as the SRT task.

Transcranial Magnetic Stimulation

Stimulation was performed using a Magstim Rapid² module (Magstim, Whitland, UK) and a 70mm diameter figure-of-eight coil. TMS pulses were delivered offline at a rate

of 1Hz, i.e., 1 pulse per second, for 10 minutes totalling 600 pulses per site (1800 pulses per experimental session). Stimulation intensity was set at 40 percent of the maximum pulse strength and held constant across all participants. Stimulation intensity was set at this level as a result of pilot testing where 40 percent was found to be safe and comfortable for the participants. Additionally, TMS coils heat up with delivery of multiple pulses and need to be replaced when they become too hot. Through pilot testing, it was found that 40 percent intensity was low enough to prevent the coils from overheating too quickly and thus allowing the coil to last the full 600 pulses without the need to switch coils. Motor thresholds were not used as their applicability to non-motor regions is yet to be fully established (Stewart, Walsh, & Rothwell, 2001; Stokes et al., 2013). All TMS parameters were well within established international safety limits (Rossi, Hallett, Rossini, & Pascual-Leone, 2009; Rossini et al., 2015)

The effects of 1Hz offline rTMS have been extensively investigated in the human motor cortex (Chen et al., 1997), with results showing a suppression of motor excitability for up to 15 minutes after 10 minutes of stimulation (Romero, Ansel, Sparing, Gangitano, & Pascual-Leone, 2002). Therefore, all participants received 10 minutes of offline rTMS at a rate of 1Hz followed by a five-minute rest period. At five minutes post-TMS participants completed a single speech reception threshold task and two gap detection threshold tasks. This test period was then followed by a rest period of at least 15 minutes where no testing or stimulation was administered to ensure enough time for the previously applied cortical suppression to abate before stimulating another site (Chen et al., 1997; Romero et al., 2002). The amount of time that had elapsed from the end of TMS to the end of each 10-minute testing session was recorded to ensure that all participants completed the three tasks before the 15-minute post-TMS time period had passed (results of a one-way repeated measures ANOVA revealed no significant difference between sites in the length of time that elapsed between the end of rTMS and completion of all tests: $F(3, 45)=1.016, p=0.394$).

Before the experiment, began all participants received three to four trains of pulses per site, to ensure they were comfortable with the stimulation parameters. During this demonstration all participants used an earplug (3M E.A.R., 36dB attenuation) in the ear ipsilateral to the site of stimulation to attenuate the sound of the

coil discharge and avoid damage to the ear (Counter, Borg, & Lofqvist, 1991). During the main experiment, magnetically shielded ER1 Etymotic earphones were used bilaterally to both deliver the auditory stimuli and attenuate the sound of the coil discharge and avoid damage to the ear.

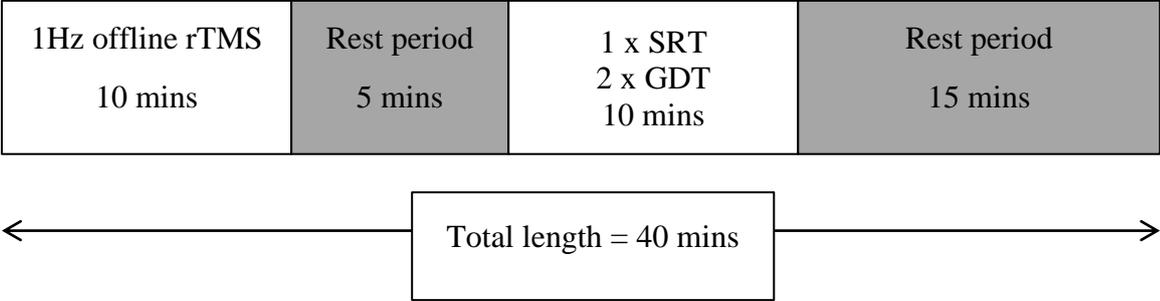


Figure 5 - Illustration of the order of events and associated timings for one cortical site in Experiment Three.

Selection of TMS sites

The experimental sites for this experiment were taken from Adank (2012) who conducted an Activation Likelihood Estimation (ALE) meta-analysis of 57 fMRI and PET studies that contrasted intelligible with less intelligible or unintelligible speech stimuli. ALE is used to establish the degree of overlap between coordinates taken from different neuroimaging papers. Across all 57 studies the site with the highest ALE score and therefore the site with the most observed activation across studies was the left superior temporal sulcus with MNI coordinates of $x = -60, y = -12, z = -6$. A less active homologous cluster was found in the right STS $x = 62, y = -8, z = -10$. These two sets of coordinates were used as guides for placement of the TMS coil. In some participants, however, these coordinates did not match up to the superior temporal sulcus, as a result adjustments of the coordinates were made on an individual participant-by-participant basis to the location of target sites to ensure consistent stimulation of the STS across all participants. However, on average, the adjustments in MNI space equated to a shift of less than 5 millimetres, which is less than the predicted spatial resolution of TMS (~5-20mm), it is therefore likely that both the original and adjusted cortical location were directly stimulated by the TMS pulses (see Table 14 for average group coordinates across experiments). These coordinates relate

to the anterior STS based on the classification of Evans (2012) and the mid STS based on the classification of Okada et al. (2010). Vertex was used as a control site in this experiment and was identified as the highest point of the skull in the midsagittal plane.

Table 14 - Average group co-ordinates across experiments.

Experiment	Left STS			Right STS		
	x	y	z	x	y	z
Three	-60.21	-12.17	-3.07	60.08	-8	-4.45
Four	-59.38	-12	-4.17	60.17	-8	-6.37
Five	-59.32	-12.36	-4.11	59.60	-8.26	-4.93
Six	-59.93	-12.35	-5.85	61.32	-8	-9.64
Seven	-60	-12	-6	62	-8	-10

Notes:

MRI scanning

Participants came to the Birkbeck-UCL Neuroimaging Centre (BUCNI) to get a T1-weighted structural magnetic resonance imaging scan. [FLASH sequence, repetition time (TR) = 12ms, Echo time (TE) = 5.6ms, flip angle = 19°, resolution 1mm x 1mm x 1mm]. Immediately after the scanning session each MRI was visually inspected for the potential presence of neuroanatomical abnormalities. If no anomalies were found the individual MRI slices were processed to create one composite image and rotated to match the orientation of the MNI-152 template brain. The participant was then invited back for the TMS session, where the scan was used in conjunction with BrainSight frameless stereotaxy (Rogue Research, Montreal, Canada). BrainSight uses an infrared camera and tracking system and displays the specific location and orientation of the TMS coil in real time on the individual participant's MRI ensuring accurate and consistent stimulation of the target and control site. If an anomaly was observed on a structural scan then the participant was referred to their general medical practitioner for follow up tests and excluded from the experiment (one occurrence in experiment five).

Data Analysis

The dependent variable for the SRT task in this experiment is the average SNR level at which reversals occurred across the 30 test sentences per condition. A one-way

repeated-measures analysis of variance (ANOVA) was conducted to investigate the effect of TMS condition on SRT performance. Additionally, the dependent variable for the GDT test is the average length in milliseconds of the gap for the final four reversals in a test, a separate one way repeated measures ANOVA was conducted for the GDT data.

In order to investigate the strength of evidence for the alternative and null hypothesis, a Bayes Factor (B) is reported for all relevant tests. Hypothesis testing dictates the creation of an alternative hypothesis (H_1), which states that a particular event will occur, as well as the corresponding null hypothesis (H_0), which states that no difference will occur. Inferential statistics are then designed to assess how much evidence is provided to support H_0 from the provided data. The main issue with this is that the associated p-value really can only make the weak distinction between evidence for H_0 or evidence for H_1 , a p-value cannot however make the third distinction of whether or not the data provide insufficient evidence to support either H_0 or H_1 . Bayes factors have the advantage of providing an insight into this middle category. Following Jeffreys (1961), a B value greater than 3 indicates substantial evidence for H_1 over H_0 . In contrast a B value of less than 0.3 indicates substantial evidence for H_0 over H_1 . Finally, a value between 0.3 and 3 indicates data insensitivity, i.e., there is insufficient evidence to preferentially support either H_0 or H_1 .

In order to establish the Bayes factor, both the H_0 and H_1 must be modelled. The model of H_0 by definition would predict zero difference between groups and will therefore be modelled as such in all subsequent analyses. The modelling of H_1 is dependent on prior knowledge and expectation of results. However, as stated above, no previous study has investigated whether or not TMS is capable of disrupting cortical auditory processing enough to shift speech reception thresholds. Therefore, in the absence of a directly comparable literature, in the following analyses H_1 was modelled on the results of Peters, Moore, and Baer (1998). This study found an SNR difference of 3.2 dB between a group of young normal hearing participants and a group of young hearing-impaired participants. The hearing impairment was the result of sensorineural deficits causing an average hearing threshold above 20dB at octave frequencies from 125 to 8000Hz). Whilst TMS is being used to interfere with central cortical/auditory processing in this study as opposed to peripheral auditory processing, the ultimate aim

would be to disrupt cortical processing to the same extent that it would produce deficits akin to perceptual deficits encountered as a result of peripheral hearing loss.

Results

Speech Reception Threshold

A repeated-measures ANOVA was conducted to investigate whether the application of TMS to different anatomical landmarks (no TMS; vertex; left STS; right STS) produced differential effects on participants' ability to perceive speech in noise. The results of this analysis were non-significant $F(3, 45)=0.928$, $p=0.43$, $\eta^2=0.05$, $B_{10}=0.21$ indicating substantial evidence in favour of the null hypothesis.

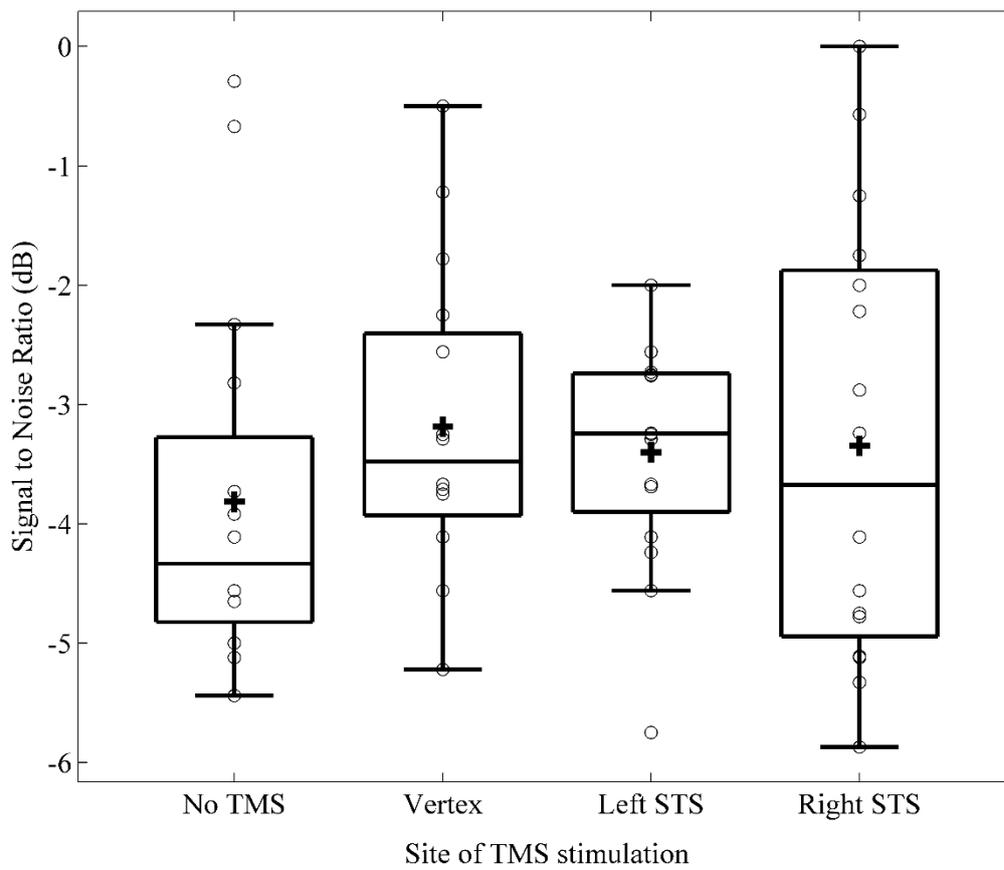


Figure 6 - Boxplots displaying SRT performance across the different TMS stimulation sites for Experiment Three.

Table 15 - Summary of group descriptive statistics for SRT task across TMS condition.

	No TMS	Vertex	Left STS	Right STS
Mean	-3.81	-3.18	-3.40	-3.34
Std.Dev.	1.53	1.24	0.92	1.85
Min	-5.44	-5.22	-5.75	-5.87
Max	-0.29	-0.50	-2.00	0.00

Notes: All values represent Signal to Noise ratios in decibels.

Gap Detection Threshold

A repeated measures ANOVA was conducted to investigate whether the application of TMS to different anatomical landmarks (no TMS; vertex; left STS; right STS) produced differential effects on the length of gap that participants could reliably detect in a burst of white noise. The results of this analysis were non-significant $F(3,$

45)=0.14, $p=0.93$, $\eta^2=0.01$, $B_{10}=0.09$ indicating substantial evidence in favour of the null hypothesis.

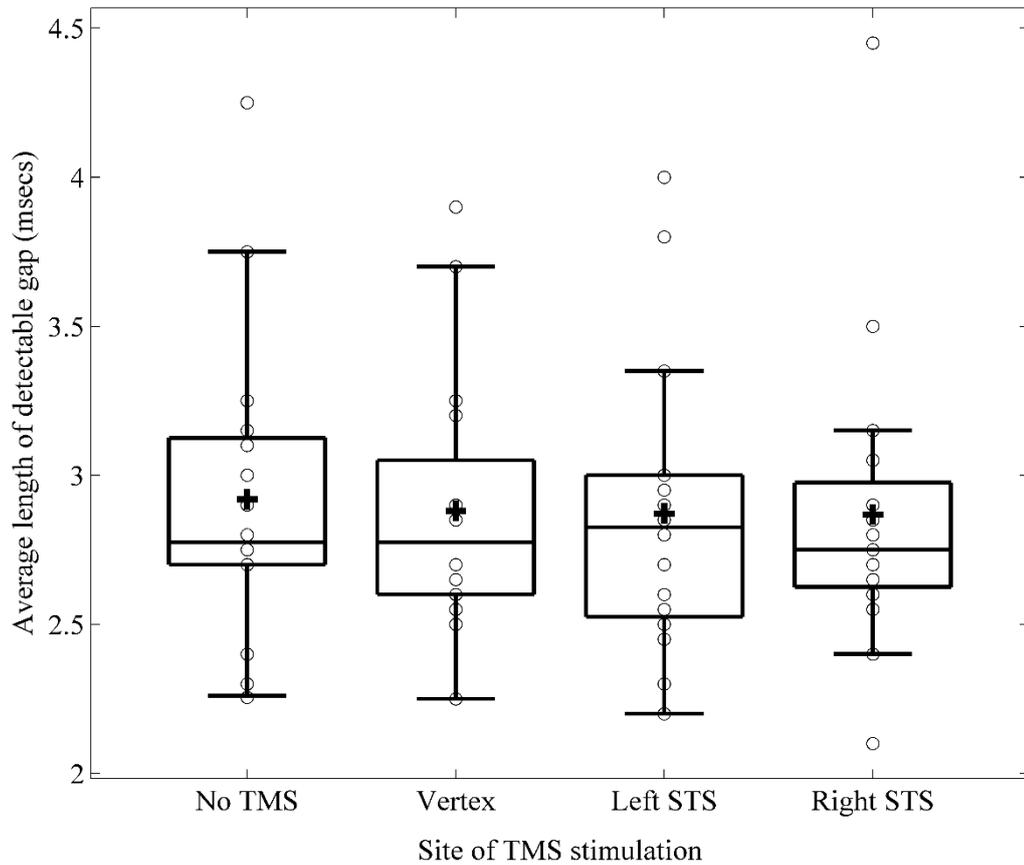


Figure 7 - Boxplots displaying GDT performance across the different TMS stimulation sites for Experiment Three.

Table 16 - Summary of group descriptive statistics for GDT task across TMS condition.

	No TMS	Vertex	Left STS	Right STS
Mean	2.91	2.88	2.87	2.86
S.D.	0.51	0.43	0.49	0.52
Min	2.26	2.25	2.20	2.10
Max	4.25	3.90	4.00	4.45

Notes: All values represent length of perceived gap in milliseconds.

Interim discussion

The current results suggest that application of offline 1Hz rTMS did not impair participants' ability to perceive either intelligible speech or noticeable gaps in noise. A consistent difference between the published research and this experiment however is the intensity of the TMS pulse (in terms of maximum output of the capacitor). There are generally two approaches used to set the intensity of the TMS pulse. The first option involves a degree of trial and error within pilot testing to find an intensity that is high enough to be effective whilst remaining low enough to be both safe and comfortable for the participants. This is the approach that was adopted for experiment three, with stimulation intensity set at 40 percent of maximum output and held constant across all participants. The second convention for setting TMS intensity is based on each individual participants' motor threshold. This is defined as the lowest level of intensity that produces a visible contraction in the hand of the participant or a noticeable deviation on an electromyographic trace, known as a motor evoked potential (MEP). The TMS intensity is then generally set at 110-120 percent of the individual's motor threshold, this approach is more common within the literature. As a result past studies comparable to this research have used an average intensity of 61 percent (see Table 13 for the average intensity used in all comparable studies), which is noticeably higher than the 40 percent intensity used in the studies presented thus far. Whilst overstimulation increases the risk of negative side effects from TMS (most notably seizure) and reduces the focality of stimulation. Under stimulation reduces the chances of finding an experimental effect. It is possible therefore that the intensity of TMS used in experiment three was too low to cause the level of cortical disruption required to reliably impair performance across all experimental conditions. This issue was addressed in experiment four where each individual participants' motor threshold level was used to guide the strength of TMS pulse for that particular subject.

Experiment Four

Methods

Participants

Sixteen different participants took part in this experiment (mean age 21yrs 1mth \pm 3.16; range 18-28; 11 females). All participants met the same eligibility criteria as outlined in the previous experiment (i.e. native British English, right-handed, with no reported history of speech, language, neurological or psychiatric disorder) and were paid for their participation.

Transcranial Magnetic Stimulation

Stimulation was performed using a Magstim Rapid² module and a 70mm diameter figure-of-eight coil (Magstim, Whitland, UK). The TMS intensity for this experiment was set at 120 percent of each individual participants' active motor threshold (aMT). To establish aMT, electromyographic (EMG) activity was recorded from the first dorsal interosseous (FDI) muscle in the right hand of each participant using surface electrodes (Ag/AgCl; 10mm diameter). The hand region of the left primary motor cortex was anatomically identified according to the appearance in the axial plane of an omega or epsilon type structure in the precentral gyrus (Yousry et al., 1997) using each individual participant's previously obtained MRI. Where necessary this location was adjusted if reliable hand MEPs were not found. Individual participant aMT was defined as the lowest TMS intensity that would reliably elicit 5 out of 10 MEP's, from the FDI muscle, with an amplitude of at least 50 μ V (Groppa et al., 2012; Rossini et al., 1994; Rossini et al., 2015). Using these criteria, the average active motor threshold was 47.03 percent (range 25-60) resulting in an average stimulation intensity of 56.44 percent of maximum stimulator output (range 30-72). Anything greater than 55 percent of maximum stimulator output at a rate of 1Hz causes the stimulating coil to overheat in less than ten minutes and thus the need to switch to another coil mid-session. Therefore, in order to standardise procedures across all participants two, six-minute blocks of 1Hz offline rTMS were conducted consecutively with a change of coil at the end of each block (360 pulses per block; 720 pulses total). This protocol was used

irrespective of aMT and subsequent TMS intensity (i.e., above or below 55 percent) to ensure consistency across participants. The order of TMS coil use was counterbalanced.

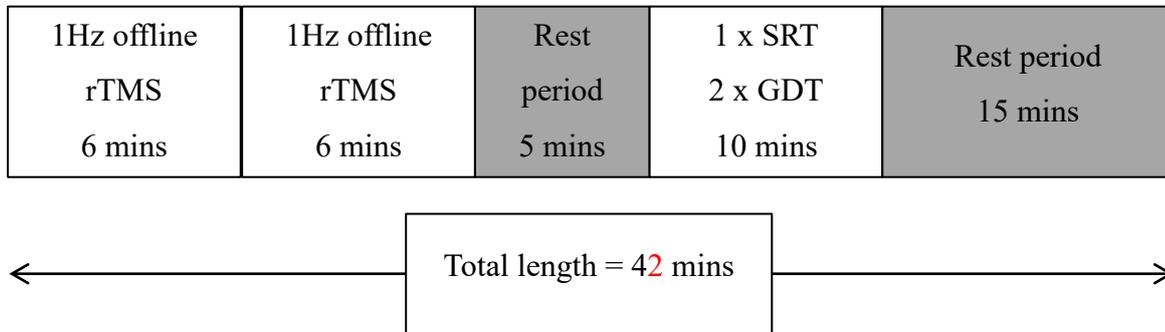


Figure 8 - Illustration of the order of events and associated timings for one cortical site in Experiment Four.

In experiment three the control condition (vertex) results were more comparable to the right and left STS conditions rather than the baseline no TMS condition suggesting a general TMS effect as opposed to a site-specific impairment. Therefore, in the following experiment, the control site was changed to the occipital pole in an attempt to uncover the causes of the impairment in performance after vertex stimulation in experiment three. Occipital pole was located as being 20mm superior and 10mm lateral of the inion, as in previous TMS studies (Krieger-Redwood et al., 2013). The direction of the lateral movement (left or right) was again counterbalanced.

Procedure

The task again consisted of a single speech reception threshold using the same four lists of sentences used in experiment three (see Appendix J) and two gap detection threshold tests. The SRT is designed to find the approximate location of the participants’ threshold as quickly as possible before reducing to a smaller step size in an attempt to maximize the number of trials that are close to the threshold level. In experiment three the final step size was 4dB, given the subtle differences in performance that were being investigated, it is possible that any potential difference in

performance across conditions were missed due to the large step change. Therefore, in the current experiment the starting SNR level was changed from +20dB to +10dB and all final step size changes were +/-2dB. By making the starting SNR value lower it reduces the number of trials required to locate the potential threshold. By making the final step size smaller it should result in a more accurate judgment of each individual threshold level potentially allowing for a clearer distinction between the conditions.

Results

Speech Reception Threshold

A repeated measures ANOVA was conducted to investigate whether the application of TMS to different anatomical landmarks (no TMS; Occipital pole; left STS; right STS) would produce differential effects on participants' ability to perceive speech in noise. A non-significant main effect of TMS condition was found $F(3,45)=0.99$, $p=0.4$, $\eta^2=0.06$, $B_{10}=0.23$ indicating substantial evidence in favour of the null hypothesis. These results suggest that participants' performance was not affected by the application of 1Hz rTMS at 120 percent of their active motor threshold.

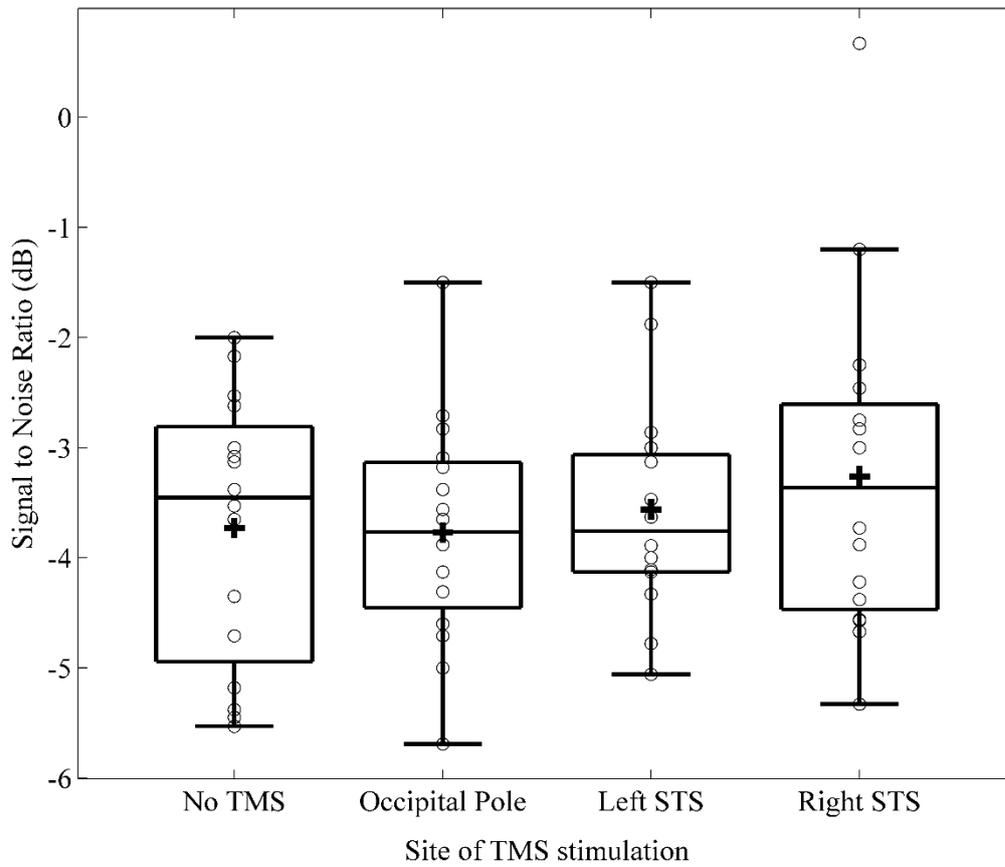


Figure 9 - Boxplots displaying SRT performance across the different TMS stimulation sites for Experiment Four.

Table 17 - Summary of group descriptive statistics for SRT task across TMS condition.

	No TMS	Occipital Pole	Left STS	Right STS
Mean	-3.73	-3.77	-3.56	-3.26
S.D.	1.209	1.01	0.96	1.51
Min	-5.53	-5.69	-5.06	-5.33
Max	-2.00	-1.50	-1.50	0.67

Notes: All values represent Signal to Noise ratios in decibels.

Gap Detection Threshold

A repeated measures ANOVA was conducted to investigate whether the application of TMS to different anatomical landmarks (no TMS; vertex; left STS; right STS) would produce differential effects on the length of gap that participants could reliably detect in a burst of white noise. The results of this analysis were also non-significant $F(3, 45)=0.69, p=0.56, \eta^2=0.04, B_{10}=0.17$ indicating substantial evidence in favour of the null hypothesis.

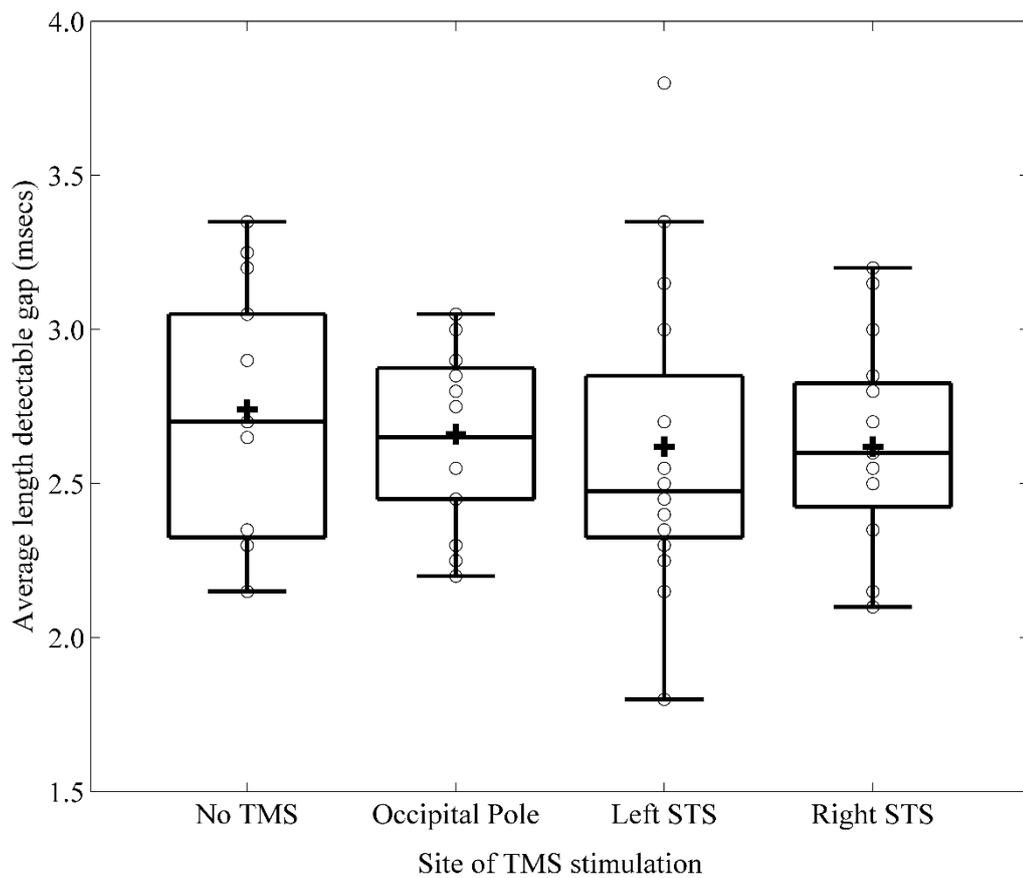


Figure 10 - Boxplots displaying GDT performance across the different TMS stimulation sites for Experiment Four.

Table 18 - Summary of group descriptive statistics for GDT task across TMS condition.

	No TMS	Vertex	Left STS	Right STS
Mean	2.74	2.65	2.61	2.61
S.D.	0.38	0.27	0.49	0.33
Min	2.15	2.20	1.80	2.10
Max	3.35	3.05	3.80	3.20

Notes: All values represent length of perceived gap in milliseconds.

Interim discussion

The results of experiment four show no main effect of stimulation site for either the speech reception or gap detection threshold tests. These results are in keeping with the results of experiment three that also found no main effect of 1Hz offline rTMS. It is interesting to note that despite the higher intensity of TMS being used in this experiment, the differences between conditions are smaller compared to the previous experiment. In itself this result highlights one of the key issues in TMS research, i.e., what is the best way to optimise stimulation intensity (Kaminski, Korb, Villringer, & Ott, 2011; Stokes et al., 2013). Overstimulation increases the risk of negative side effects (i.e., seizure) whilst under stimulation reduces the chances of finding an experimental effect. Due to the inability to functionally localise regions outside of motor and visual cortex using TMS, the use of motor thresholds as a proxy of overall cortical excitability/responsiveness to TMS has become a standard procedure, although functional localisers have been established for the supramarginal gyrus (Sliwinska, James, & Devlin, 2015; Sliwinska, Khadilkar, Campbell-Ratcliffe, Quevenco, & Devlin, 2012). Yet the relationship between the excitability of the motor cortex and other cortical structures is still a long way from being fully understood. Stokes et al. (2013) suggest that the use of the motor threshold as a calibrator for intensity of non-motor regions is valid as long as the difference in cortical depth between the motor cortex and the target cortical site is taken into account. The authors put forward the following equation to assist in the calibration between sites:

$$\text{AdjMT\%} = g \times (\text{DSiteX} - \text{DM1}) \times \text{MT}_{\text{M1}}$$

Where g = a general scaling coefficient which Stokes et al. (2013) recommend to be 2.7, reflecting a change of 2.7 percent in stimulation intensity per millimetre difference in depth between M1 and the target site; $DSiteX$ = the cortical depth of the stimulation site of interest; $DM1$ = the cortical depth of the motor cortex and MT_{M1} reflects the observed motor threshold. Using the above equation in a retrospective analysis of the cortical depth of the motor and superior temporal sulcus sites across participants suggests that the ideal TMS intensity for this experiment (based on motor threshold excitability) would have been 33 percent (more accurately 32.62). This equates to a TMS intensity that is 23 percent lower than the level used in experiment four (average intensity = 56.44), but only 7 percent lower than the intensity used in experiment three (intensity held at 40 percent for all participants). It is interesting to note that experiment three resulted in greater differences (although equally non-significant) between TMS conditions suggesting that simply increasing the TMS intensity is not necessarily the most effective way to find an effect. It is thus possible that the frequency of the pulses are the key determinant on whether or not a site responds to stimulation as opposed to the intensity per se. To investigate this further, in experiment five an offline repetitive TMS paradigm was again used but pulses were delivered in a much shorter and more compressed time frame at a rate of 10Hz starting 1500 milliseconds before each sentence. The onset of each sentence coincided with the offset of each train of TMS pulses. Stimulation of cortical regions immediately before stimuli presentation has been shown to be effective in disrupting visual processing (de Graaf, Cornelissen, Jacobs, & Sack, 2011).

Experiment Five

Methods

Participants

Sixteen different participants took part in this study (mean age 23yrs 2mths \pm 3.25; range 18-30; 8 female). All participants met the previously outlined eligibility criteria (i.e. native British English, right-handed, with no reported history of speech, language, neurological or psychiatric disorder) and were paid for their participation.

Transcranial Magnetic Stimulation

Stimulation was performed using a Magstim Rapid² module and 70mm diameter figure-of-eight coil (Magstim, Whitland, UK). Stimulation intensity was returned to 40 percent of maximum stimulator output and held constant across all participants. Stimulation intensity was set at this level based on the results of experiment three and four wherein 40 percent was found to be more comfortable for the participants (and possibly more effective) compared to the use of aMT.

As outlined above this study again used an offline repetitive TMS paradigm but adopted a different pulse frequency. In the current study pulses were delivered offline at a rate of 10Hz for 1.5 seconds (15 pulses per trial) starting 1500 milliseconds before each sentence. Timings were such that the onset of each sentence coincided with the offset of each train of TMS pulses. All participants received three to four trains of pulses per site prior to the start of the experiment to ensure they were comfortable with the stimulation parameters.

Whilst this stimulation protocol would still be considered offline, given its proximity to sentence onset, the control site in this study was again changed. In this study, the left lateral occipital complex (LOC) was used. The LOC was adopted as a control site because stimulating this region of cortex produces the same distracting side effects (e.g. facial twitches) as with stimulation of the left or right STS, therefore the LOC was considered to be a more comparable and therefore appropriate control site for this experiment.

Procedure

The same four lists of sentences from experiments three and four were used in this experiment (see Appendix J), with all sentences delivered binaurally via magnetically shielded Etymotic ER-1 earphones. Speech reception threshold parameters included final reversal steps sizes of 4dB, with the initial starting SNR remaining at +10dB (as in experiment three). The gap detection task was not used within this experiment due to the consistent lack of a TMS related behavioural modulation in the previous two

experiments. This provided the opportunity to halve the required number of pulses (from 3600 to 1800) in this experiment to keep them consistent with experiments three and four thus making the experiment safer and more comfortable for the participants.

Results

A repeated measures ANOVA was conducted to investigate whether the application of TMS to different anatomical landmarks (no TMS; LOC; left STS; right STS) would produce differential effects on participants' ability to perceive speech in noise. A non-significant main effect of TMS condition was found $F(3,45)=0.64$, $p=0.59$, $\eta^2=0.04$, $B_{10}=0.16$ indicating substantial evidence in favour of the null hypothesis.

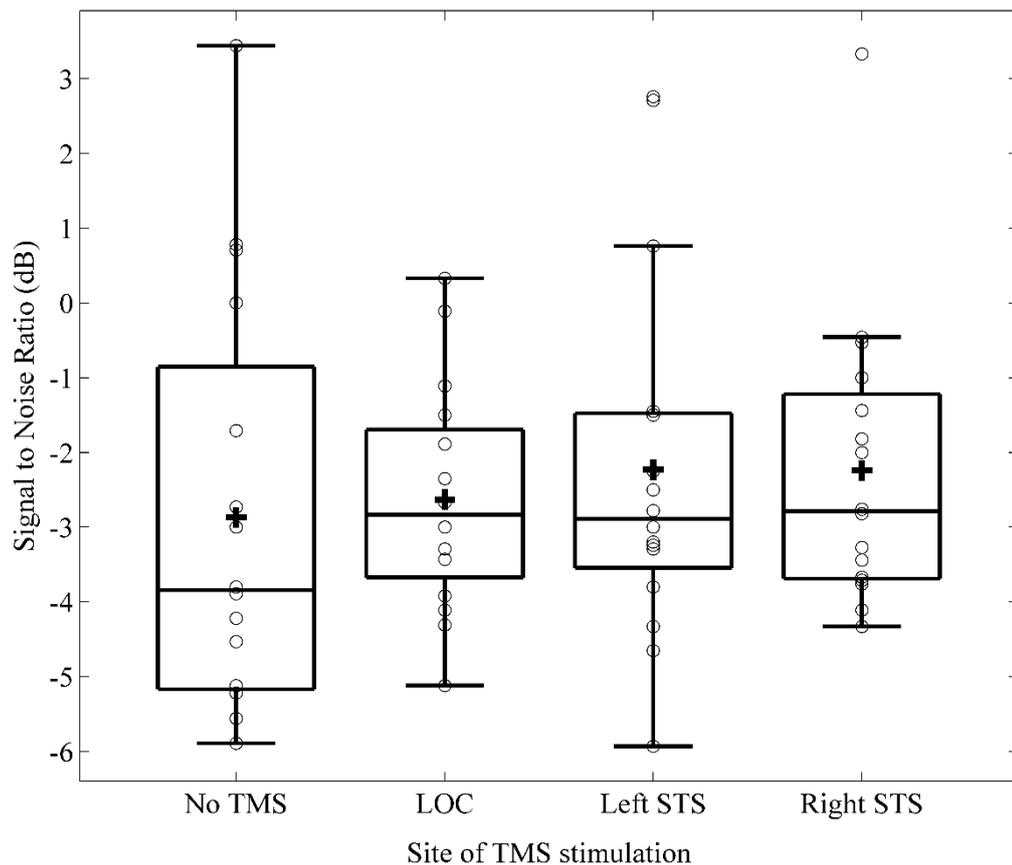


Figure 11 - Boxplots displaying SRT performance across the different TMS stimulation sites for Experiment Five.

Table 19 - Summary of group descriptive statistics for SRT task across TMS condition.

	No TMS	LOC	Left STS	Right STS
Mean	-2.87	-2.63	-2.23	-2.23
S.D.	2.76	1.49	2.44	1.94
Min	-5.89	-5.12	-5.93	-4.33
Max	3.44	0.33	2.76	3.33

Notes: All values represent Signal to Noise ratios in decibels.

Interim Discussion

In the current experiment the frequency of TMS pulse delivery was changed from 1 pulse per second (i.e., 1Hz) to 10 pulses per second (i.e., 10Hz), once again however no significant effect of offline rTMS was found on participants' ability to perceive speech in noise. One of the final possible causes for the lack of a significant TMS effect on performance could be due to the focal effects of TMS and the degeneracy of the speech perceptual system. Price and Friston (2002) argue that most cognitive models indicate that there are multiple ways of completing the same cognitive task. As a result, whilst most tasks have a set of dominant cortical regions responsible for the requisite processing under optimal conditions, these dominant cortical areas are not the only regions capable of conducting the required actions. Therefore, when damage occurs to the dominant default regions, one of the undamaged subsidiary regions takes over the responsibility. As a result, behavioural effects are not always found following known cortical damage, however this should not be interpreted as the cortical damage having no effect on functioning.

This plastic compensatory process was shown by Andoh and Paus (2011) who combined 1Hz offline rTMS with functional imaging and showed a task-related increase in activation in the homologue areas contralateral to the site of stimulation. This result was ascribed to the brain compensating for the TMS induced disruption to one hemisphere by drawing in additional resources from the opposite hemisphere. All of the experiments presented thus far have used offline rTMS with a separation in time between application of TMS and completion of the task, it is therefore possible that the lack of a significant TMS effect is due to neural mechanisms compensating for the induced subtle modulation before the participants are required to complete the task. Previous papers have suggested that such compensation could occur in less than four minutes post stimulation (O'Shea, Johansen-Berg, Trief, Göbel, & Rushworth, 2007), it is therefore especially likely that the five minute post stimulation period that was used in both experiment three and four could have left too much time for the compensatory neural mechanisms to overcome any effects of the offline rTMS. Despite no significant effect occurring within any of the individual experiments, overall performance in experiment five is reduced (by roughly a decibel) in comparison to experiments three and four, this could be evidence of partial but not full

compensation for the TMS induced modulation. As a result, in experiment six an online rTMS paradigm was utilised to maximise cortical modulation within the superior temporal sulcus whilst minimising the possibility for the intra- or interhemispheric compensation to placate the behavioural effects of this modulation.

Experiment Six

Methods

Participants

Sixteen participants (mean age 23yrs 4mths \pm 6.94; range 18-41; 11 females) were recruited for this experiment. All participants met the previously outlined eligibility criteria (i.e. native British English, right-handed, with no reported history of speech, language, neurological or psychiatric disorder) and were paid for their participation. Two of the participants in this study also took part in experiment three.

Test of Etymotic ER1 earphones

Prior to the main experiment a test of the attenuation capabilities of the ER1 earphones was conducted to investigate whether the acoustic click of the TMS coil would interfere with the main experimental task. A B&K 4157 coupler was used (Brüel & Kjær sound and vibration measurement, Nærum, Denmark) with the output connected to the left channel of a Scarlett 2i2 USB interface (Focusrite Audio Engineering Ltd, High Wycombe, U.K.). The Scarlett 2i2 USB interface was adjusted such that with the ER1 not inserted into the coupler and the Magstim rapid² (Magstim, Carmarthenshire, U.K.) module running at 10Hz, 100 percent maximum intensity (i.e., the 4157 responding to the acoustic click from the TMS coil) the recorded level was about six decibels below overload. The ER1 inputs were connected to 50ohm terminators, and only the right channel ER1 was used for the measurements, which were recorded using cooledit 96 (Adobe systems, Inc., San Jose, USA) at a sampling rate of 44.1kHz, 16 bit.

A 70mm diameter figure-of-eight TMS coil was held approximately 30cm above the ER1 shielded transducer box. With the ER1 not inserted into the coupler the Magstim rapid2 module was run at a rate of 10Hz, 100 percent of maximum pulse strength. Under these conditions, the acoustic click associated with firing the TMS coil was recorded at a level of 81.9 dB SPL. Then, in order to assess the acoustic leakage through the foam insert of the ER1 earphones, with the TMS coil held in the same position, the ER1 was inserted into the B&K coupler and the Rapid2 module was again run at 10Hz, 100 percent maximum stimulator output. Under these conditions, the acoustic click of the TMS coil was recorded at 37.8 dB SPL, inferring an attenuation of 44.1 dB, resulting in a level of background noise that was believed to be low enough to not impact upon the main experimental task. This was confirmed anecdotally when all participants reported being able to comfortably hear the sentences over the noise of the TMS pulses.

Transcranial Magnetic Stimulation

Stimulation was performed using a Magstim Rapid² and a 70mm figure-of-eight coil (Magstim, Whitland, UK). Pulses were delivered online (i.e., at the same time that the sentences were presented) at a rate of 10Hz for 2500 milliseconds, starting 500 milliseconds before each sentence began and continuing until the sentence had finished (25 pulses per trial). Stimulation intensity was set at 40 percent of the maximum pulse strength and held constant across all participants.

Before the experiment began all participants received three to four trains of pulses per site, to ensure they were comfortable with the stimulation parameters. During this demonstration all participants used an earplug (3M E.A.R., 36dB attenuation) in the ear ipsilateral to the site of stimulation to attenuate the sound of the coil discharge and avoid damage to the ear (Counter et al., 1991). During the main experiment, magnetically shielded ER1 Etymotic earphones were used bilaterally to both deliver the auditory stimuli and attenuate the sound of coil discharge.

Procedure

Four different lists 30 sentences were created from the same pre-recorded set of IEEE sentences (see Appendix K). The same male speaker of standard southern British English read all sentences in a sound attenuated room. Audio digitising was performed at 44.1kHz and 16 bits. The beginning and end of each sentence was trimmed to zero crossings as closely as possible to the onset/offset of the initial and final speech sounds. The sentences were then peak-normalised to 99 percent of maximum amplitude and scaled to 70dB SPL using Praat (Boersma & Weenink, 2011). The sentence lists were changed from previous experiments due to the recruitment of participants that had partaken in previous experiments.

Results

A repeated-measures analysis of variance (ANOVA) was conducted to investigate whether the application of TMS to different anatomical landmarks (no TMS; vertex; left STS; right STS) would produce differential effects on participants' ability to perceive speech in noise. Due to the functional relevance of bilateral superior temporal sulcus in speech perception, overall thresholds were expected to be higher, representing poorer performance, after separate application of TMS to both the left and right STS conditions relative to the no TMS and vertex control conditions. A significant main effect of TMS condition was found $F(3,45)=10.47, p<0.001, \eta^2=0.41, B = 977.84$ indicating substantial evidence in favour of the alternative hypothesis which stated that the performance of a speech reception threshold task will differ depending on site of stimulation. Post-hoc paired samples *t*-tests confirmed the hypothesis that stimulation of left (mean $-1.64\text{dB} \pm 1.61$) and right STS (mean $-0.99\text{dB} \pm 1.81$) impaired perception of sentences presented in noise showing significant differences between both the experimental sites and the no TMS (mean $-2.96\text{dB} \pm 1.57$) and vertex (mean $-2.81\text{dB} \pm 1.67$) stimulation conditions with substantial evidence in favour of the alternate hypothesis in all experimental site vs. control site contrasts (see Table 20 for all relevant statistics).

No difference was observed between either of the control conditions with Bayesian analyses suggesting the data provide substantial evidence in favour of the

null hypothesis (i.e., no difference will exist between these conditions). Additionally, no significant difference was also observed between the left and right STS performance, follow up Bayesian analyses suggest that the data provide insufficient evidence ($B=0.63$) to really favour a conclusion of there being no difference between these conditions.

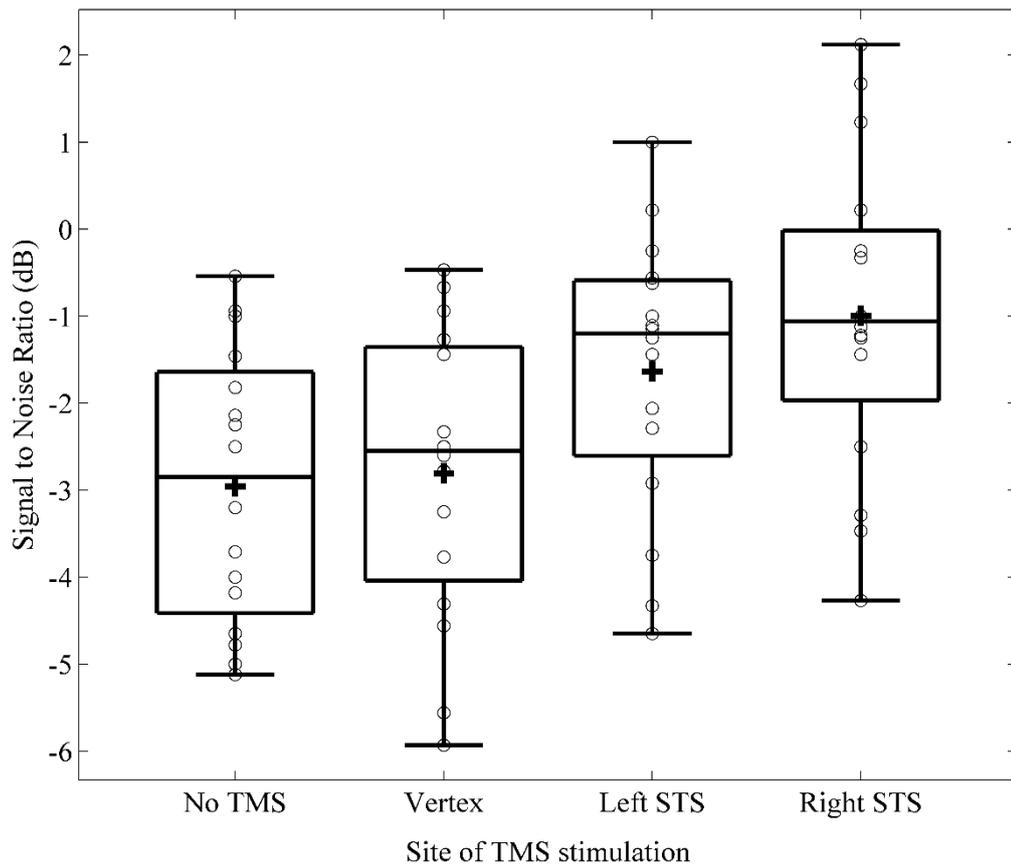


Figure 12 - Boxplots displaying SRT performance across the different TMS stimulation sites for Experiment Six.

Table 20 - Summary of the pairwise comparison statistics for experiment six.

Site A	Site B	t	p	Mean Difference	Confidence Interval	Cohen's d	Bayes _(0,3,2)
Left STS	Right STS	-1.30	0.213	-0.64	[-1.69, 0.4]	-0.325	0.63
Left STS	No TMS	3.12	0.007	1.32	[0.41, 2.22]	0.78	31.46

Left STS	Vertex	3.64	0.002	1.17	[0.48, 1.85]	0.91	143.51
Right STS	No TMS	4.57	<0.001	1.96	[1.04, 2.87]	1.14	7843.76
Right STS	Vertex	4.19	0.001	1.81	[0.89, 2.73]	1.04	1505.34
No TMS	Vertex	-0.42	0.676	-0.15	[-0.9, 0.6]	-0.105	0.16

Notes: Bonferroni corrected alpha-level = $(0.05/6) = 0.008$

Interim discussion

Speech reception thresholds were found to be elevated, reflecting poorer performance, after application of online rTMS to both left and right superior temporal sulci compared to a no TMS control condition and the TMS control site (Vertex). These results showed for the first time in this series of experiments that TMS appeared to affect auditory sentential processing within (secondary) regions of the auditory cortex. These results are important for two reasons; firstly, they provide a theoretical basis upon which more informative studies concerning the functional importance of specific regions within the auditory cortex can be developed e.g. which cortical regions are involved in adapting to distorted speech. Secondly, the equal drop in performance across the left and right STS stimulation conditions supports the notion of bilateral processing in speech perception.

TMS research paradigms usually incorporate a stimulation control site, i.e., a site that is stimulated despite its lack of functional relevance to the task/behaviour under investigation. This is to ensure that any observed changes in behaviour are caused by the intended disruption of cortical processing at the main experimental site and are not caused by general changes in attention/arousal caused by the TMS click and/or skin sensation that occur every time a TMS pulse is discharged. Whilst vertex is a cranial landmark (the highest point on the top of the skull) as opposed to a specific brain region it is commonly used as a control site in TMS research. This is mainly due to its ease of localisation as well as the lack of directly underlying neural tissue which means that TMS to this site is not expected to influence on-going task related processing (although TMS of vertex does produce some functional changes, (Jung, Bungert, Bowtell, & Jackson, 2016)). As a result, vertex provides a good baseline

against which the results from stimulation of experimental sites can be compared to investigate general attentional/arousal effects of TMS.

The lack of a behavioural effect caused by vertex stimulation in the current study suggests that the effects observed after left and right STS stimulation cannot be fully ascribed to some of the confounding attentional/arousal side effects discussed above. However, the validity of the results may be questionable due to the fact that the application of TMS can and did directly innervate the temporalis muscles of participants, which results in the twitching of facial muscles. These twitches can at times be distracting and uncomfortable (Duecker & Sack, 2015). On the basis that such facial twitching does not occur during vertex stimulation (due to lack of facial muscles running along the top of the skull) it is possible that the participants in this experiment were simply more distracted in the left and right STS stimulation conditions during stimulus presentation compared to either the control site or the participants in the previous experiments. Therefore, in order to establish the validity of the significant results found in this experiment, a follow up experiment was conducted. In this follow-up experiment, participants completed a visual discrimination threshold task under the same four TMS conditions as used in experiment six (left STS; right STS; vertex; no TMS). A visual discrimination threshold task was used, as it is a task for which the bilateral superior temporal sulcus is functionally irrelevant. Therefore, if a disruptive effect of TMS site is found for this task, it strengthens the notion that the observed effects of experiment six are due more to the confounding side effects of TMS application (e.g. facial twitches). However, if no effect of TMS is found in experiment seven it supports the notion that the results of experiment six can be attributable more to the modulation of cortical functioning than to side effects associated with online TMS.

Experiment Seven

Methods

Participants

17 participants took part in this study, all of whom met the same eligibility criteria as outlined in previous experiments and who were paid for their participation. In addition to the previously outlined eligibility criteria (i.e. native British English, right-handed, with no reported history of speech, language, neurological or psychiatric disorder), the participants' visual acuity was assessed to establish if it was within the normal range (all participants had a binocular vision rating of less than 0.1 on the LogMAR scale equating to greater than 0.8 on the decimal scale; Colenbrander (2002)) and on average participants were capable of accurately verifying 80.46 ± 11.18 written sentences in two minutes, at an average of 1397.56 ± 192.29 milliseconds per sentence (as assessed via the Speed and Capacity of Language Processing (SCOLP) test; Baddley et al. (1992) see Appendix L). One participant was excluded from the final analysis for not completing the visual discrimination task as instructed. This participant was observed to repeatedly press the response keys throughout testing even at times when responses were not expected, i.e., no stimuli were present on the screen (final analysed $n=16$; mean age 21yrs 6mths ± 2.07 ; range 18-25; 8 females).

Procedure

During the visual discrimination threshold task, each trial began with a fixation cross displayed in the centre of the screen for 500 milliseconds; followed by a blank screen for 500ms and finally two sets of letter strings and another fixation cross were presented on screen for 2000ms. Again, the fixation cross appeared in the very centre of the screen with one of the letter strings just above and one of the letter strings just below the fixation cross (see Figure 13). After the 2000ms had surpassed the screen again went blank until the next trial began (inter-trial interval = 4000ms).

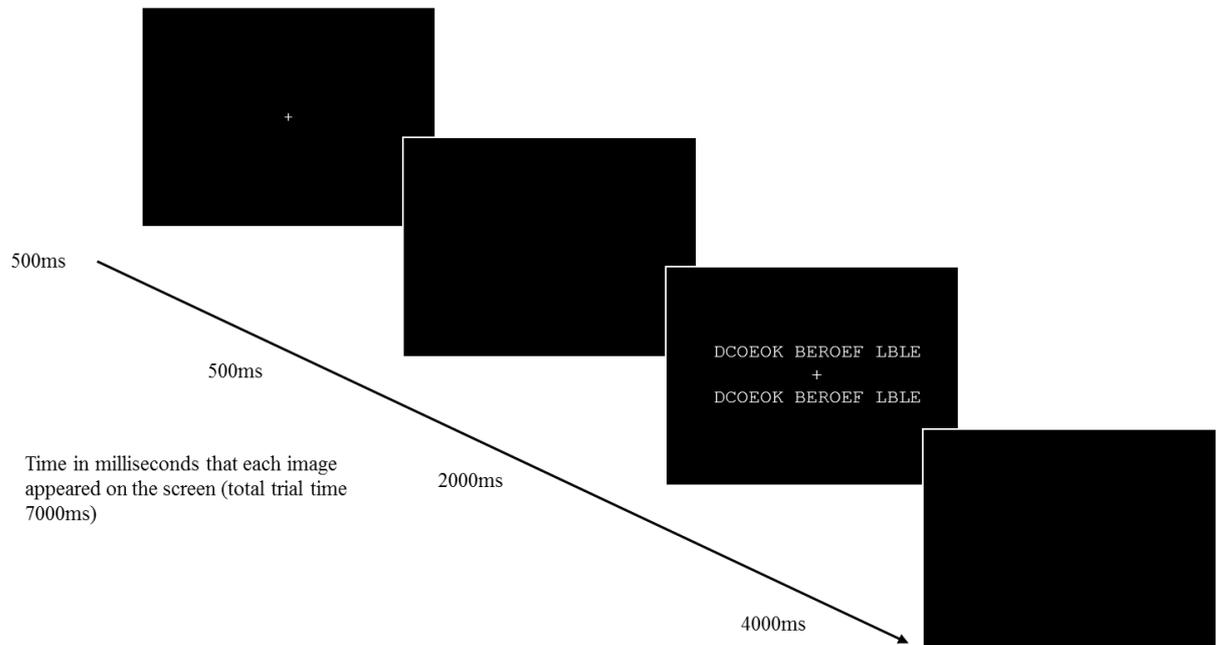


Figure 13 - Illustration of a single trial and associated timings for Experiment Seven.

The stimuli consisted of scrambled written versions of three of the five keywords used per trial in experiment six, an example of key words used in experiment six are: COOKED BEFORE BELL, in experiment seven these were visually presented as DCOEOK BEROEF LBLE. On an “identical trial” participants would simply see this letter string presented concurrently above and below the central fixation cross. On a “different trial” three of the middle letters were changed in one of the three nonsense words. The first and last letters of all nonsense words were always held constant on different trials so that matching could not rely solely on the initial and final letter. Additionally, all stimuli were presented using Courier New in font size 60. This is a fixed width font and therefore both sets of letter strings occupied the same horizontal space and thus matching had to rely on more than simple length comparisons.

The study consisted of 120 trials divided up into 30 trials per TMS condition. Of the 30 trials, 15 were identical and 15 were different. On the 15 “different trials” the change occurred five times equally across the first, second and third word. Letters were changed by simply replacing the three relevant letters with the next letter in the English alphabet, for example DCOEOK BEROEF LBLE became DCOEOK BESPPF LBLE. Nonsense letter strings were used in place of real words in order to

avoid ceiling effects in performance (see Appendix M) thus making an effect due to TMS modulation possible.

In order to make the visual discrimination threshold task as comparable to the SRT task used in experiments three-six, a staircase procedure was again adopted. In the same way that the level of the speech shaped background noise varied adaptively dependent on performance, in the current experiment the contrast level between the background and foreground (i.e., the visually presented text) was varied adaptively. On all trials, the background was black with an RGB value of [0,0,0], on the first trial the letter strings appeared with an RGB value of [0.8, 0.8, 0.8], and therefore appeared as white text on a black background. Correct discrimination resulted in an initial contrast change of +/-0.1. As a result, correct discrimination resulted in a text RGB value of [0.7,0.7,0.7] on the subsequent trial whilst incorrect discrimination would result in a text RGB value of [0.9,0.9,0.9]. This change occurred for the first 10 trials; for trials 11 to 16 contrast changes occurred in steps of 0.05; trials 17 to 25 in steps of 0.025 and 0.001 for trials 26 to 30.

As with experiments three-six, participants visual discrimination thresholds were computed by taking the mean RGB value of the letter strings for all trials where a reversal occurred (scores closer to zero represent better overall performance). Orders of stimuli list and stimulation sites were counterbalanced. All stimuli lists were pseudo-randomly ordered such that the order of presentation was different between participants but each set of three nonsense letter string ‘sentences’ only appeared once per participant. During pilot testing it was found that this task involved a large learning effect, therefore all participants completed 60 practice trials before starting the actual experimental session. The TMS stimulator and procedure were identical to those used in experiment six.

MNI-152 structural brain scan

In place of a participant specific structural scan the MNI-152 brain was used to guide placement of the TMS coil with individual MRI structural scans not obtained for any participants in the current study. In conjunction with Brainsight 2.3.5 the MNI-152 brain was adapted based on a minimum of five separate estimations of the front-, back-

, top-, left- and rightmost points on each participant's head with the MNI brain adapted to meet the measured dimensions. TMS target locations were the same as used in all previous experiments in this chapter: left superior temporal sulcus ($x = -60, y = -12, z = -6$); right superior temporal sulcus ($x = 62, y = -8, z = -10$); vertex ($x = 0, y = 0, z = 90$) and a no TMS baseline condition.

Results

A repeated-measures analysis of variance (ANOVA) was conducted to investigate whether the application of TMS to different anatomical landmarks (no TMS; vertex; left STS; right STS) produced differential effects on participants' ability to discriminate between two nonsense letter strings at variable degrees of visual contrast. A non-significant main effect of TMS condition was found $F(3,45)=1.08, p=0.367, \eta^2=0.067, B = 0.26$, indicating substantial evidence in favour of the null hypothesis which stated that the performance of a visual discrimination threshold task did not differ regardless of site of stimulation. In order to ensure that no significant differences are masked by an overall non-significant main effect, follow up post-hoc analyses were conducted without any correction for multiple comparisons. Even in these liberal circumstances all comparisons returned non-significant results (all p 's > 0.06).

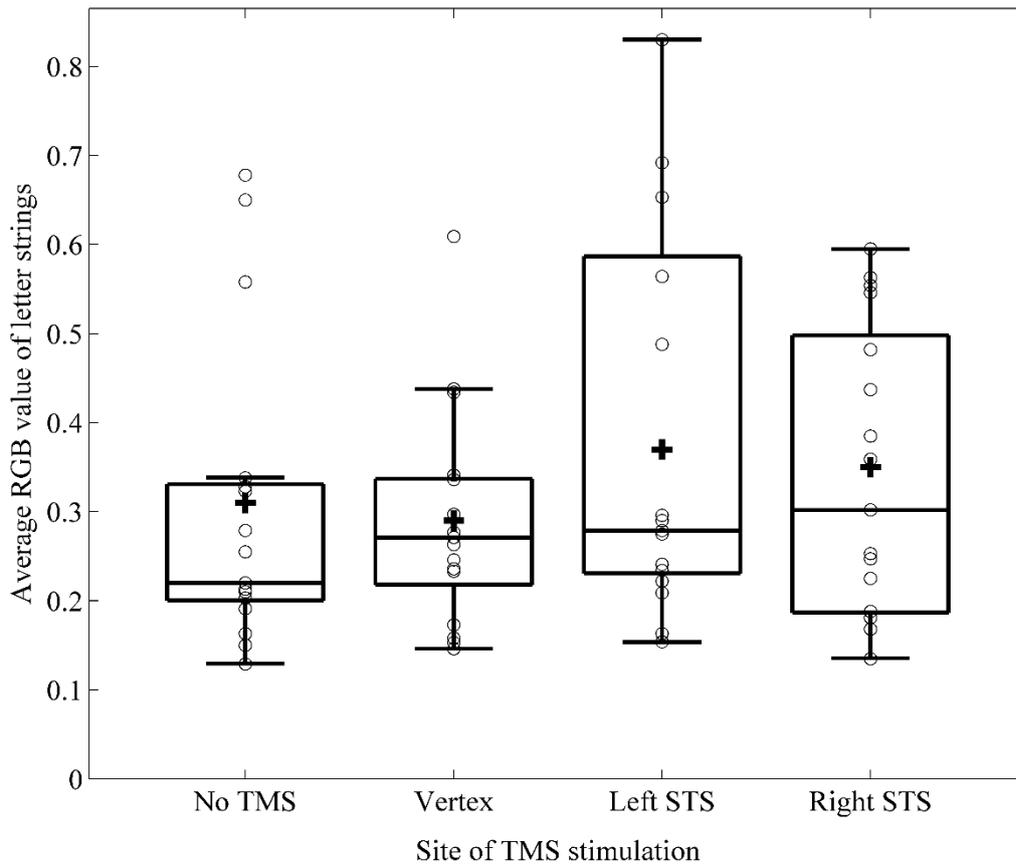


Figure 14 - Boxplots displaying Visual Discrimination threshold performance across the different TMS stimulation sites for Experiment Seven.

Table 21 - Summary of group descriptive statistics for SRT task across TMS condition.

	No TMS	Vertex	Left STS	Right STS
Mean	0.305	0.28	0.36	0.35
S.D.	0.17	0.12	0.209	0.15
Min	0.13	0.15	0.15	0.14
Max	0.68	0.61	0.83	0.59

Notes: All values represent RGB values of the letters. The RGB values of the background were [0,0,0] throughout. Numbers closer to 0 represent better performance.

Experiments Six and Seven

Results

A two-way mixed analysis of variance (ANOVA) was conducted to investigate whether the application of TMS to different anatomical landmarks (within-subjects factor with levels: no TMS; vertex; left STS; right STS) produced differential effects on two different perceptual tasks (between-subjects factor with levels: speech reception threshold vs visual discrimination threshold). A non-significant interaction between TMS condition and task was found $F(3,90)=1.68$, $p=0.18$, $\eta^2=0.04$, indicating substantial evidence in favour of the null hypothesis.

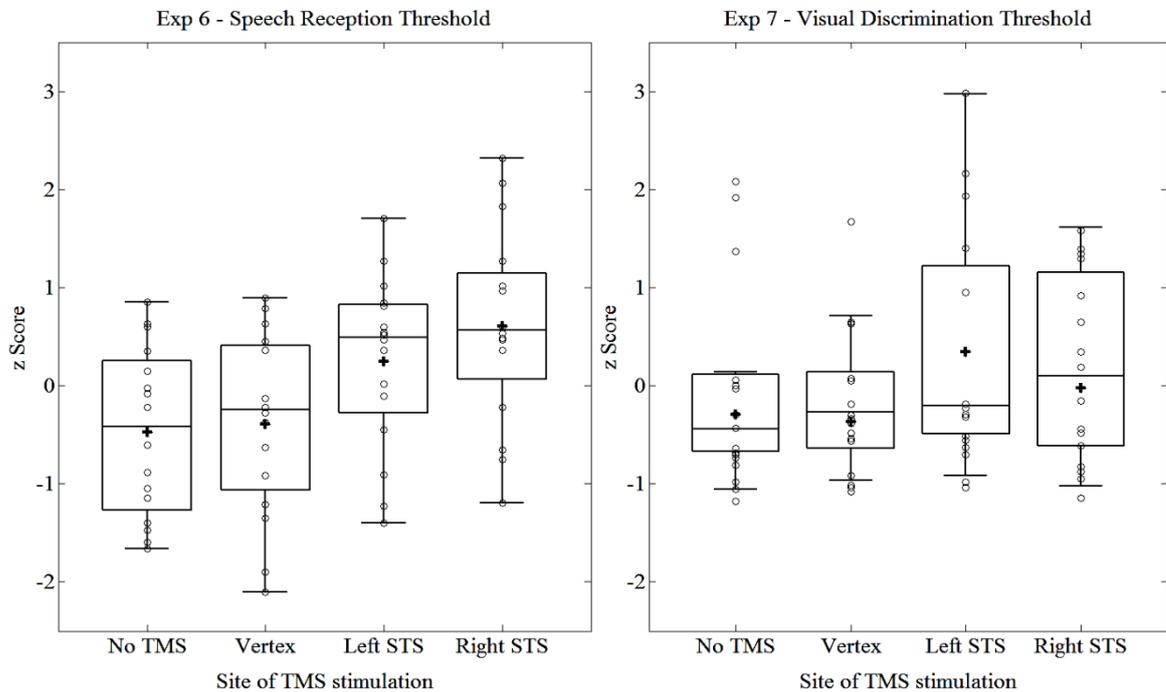


Figure 15 - Boxplots displaying z-scores across the different TMS stimulation sites for Experiments Six and Seven.

General Discussion

The present set of five experiments aimed to find the most effective transcranial magnetic stimulation protocol to non-invasively impair perception of sentences presented in noise using healthy human adults. The first three experiments all used offline repetitive TMS to modulate performance on the speech in noise task. However,

despite varying the rate, intensity and duration of pulses, no significant effect was found in any of the first three studies. As a result, in experiment six, the timing of the TMS pulses was changed from offline to online in an attempt to maximise cortical modulation whilst minimising the possibility for intra- or inter-hemispheric compensation. This change in the timing of the pulses was found to be critical in producing a significant and selective modulation of participants' ability to perceive speech in noise after both left and right STS stimulation. It was hypothesised that the effects of rTMS in experiments three, four and five were being counteracted by the compensatory processes reported by Andoh and Paus (2011) and others (Mason et al., 2014; O'Shea et al., 2007).

Lastly, in experiment seven participants completed a visual discrimination threshold task whilst rTMS was applied online at a rate of 10Hz to either their left STS; right STS; vertex (control site) or in a no TMS, control condition. The results of a one-way repeated measures analysis of variance indicated no significant difference in overall performance regardless of TMS condition. Additional Bayesian analyses showed that the data provided substantial evidence in favour of the null hypothesis. The non-significant result in experiment seven strongly suggests that participants were able to maintain enough attention despite innervation of facial musculature to complete the task in a valid way. This is critically important for the current set of experiments and future studies of speech perception that utilise TMS, as it highlights that online rTMS can be used without confounding the results, as also supported by Bestelmeyer et al. (2011). The importance here is highlighted by the results of Andoh and Paus (2011) who have shown that the application of offline rTMS results in compensatory modulations in ipsi- and contra-lateral regions of the brain to an extent that the behavioural perturbations induced through TMS can be overcome. When investigating action selection with TMS and fMRI O'Shea et al. (2007) found that these compensatory processes occur within the first four minutes after TMS induced neural modulation. Therefore, by using an online, as opposed to an offline, rTMS paradigm, experiment six afforded the opportunity to investigate the immediate impact of the disruption before any (or at least before the majority of) cortical adaptation occurred. This paradigm provided the closest possible insight into the immediate impact that neural trauma has on a task such as speech perception providing a better opportunity

to ascertain the importance of different candidate regions. Whilst we were unable to (nor did we expect to) induce a level of performance akin to clinical neurological trauma such as verbal auditory agnosia, the results of experiment six do show that both the left and right superior temporal lobes play an important role in successful speech perception.

Whilst the effect in experiment six is significant between bilateral STS and the control conditions, the overall magnitude of the effect is subtle. The just noticeable difference (JND) refers to the minimum level by which a stimulus must change before the difference is noticeable. Whilst there is still some disagreement as to the exact JND for speech embedded in noise, it is believed to be roughly two to three decibels (Killion, 2004; McShefferty, Whitmer, & Akeroyd, 2015). This suggests that for a listener to gain any benefit from noise reduction in an acoustic signal the noise would have to be reduced by a minimum of two decibels. Furthermore, Whitmer, McShefferty, and Akeroyd (2016) found the just *meaningful* difference, i.e., the minimum level at which a signal must change before listeners would be willing to change their behaviour (e.g. swapping one hearing aid for another) is roughly six decibels. In comparison, the observed difference found in experiment six of one to two decibels between the left and right STS compared to the no TMS and vertex condition could be considered negligible in a real-world setting. However, this should be considered as a general limitation of TMS as a research technique as opposed to a limitation of the current results. Whilst the level of cortical modulation in TMS studies can be enough to impair performance allowing causal inferences concerning the role of certain regions on a specific task, the impairment in performance is often reflected in very subtle changes, i.e., hundreds of milliseconds delay in response times or a few percentage points in accuracy (Silvanto & Muggleton, 2008). Therefore, an important point to consider is not the size of the effect in real world circumstances but instead whether or not a significant effect occurs in the context of the experimental design (de Graaf & Sack, 2011). In experiment six, a significant effect of rTMS was found when applied online to the left and right STS regions, and even though the effect is small in real world terms, it is theoretically important and should be considered in the context of the null effect on all control sites (vertex and no TMS) and the control task (visual discrimination) used in experiment seven.

The lack of a significant interaction between TMS site and task in experiments six and seven would suggest that the trends in the two data sets are similar and therefore it may be argued that the application of online rTMS produced an element of distraction on both tasks. However, the lack of a significant main effect of TMS site on the visual discrimination task whilst one was present on the speech discrimination task is indicative that any level of distraction was very minimal. The lack of a significant interaction is believed to be due to a lack of statistical power for a two by four mixed F-test. When deciding on the number of participants to test in the current series of experiments, the related published literature was consulted. The average number of participants across 13 related studies published prior to the commencement of experiment one is 14.54. Therefore, based on the previous literature, use of 16 participants per study in the current series of experiments appeared to be sufficient for a one-way ANOVA with four levels. However, when (post-hoc) calculating the required sample size for a two by four within-between subjects interaction using G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007) with a medium partial eta-squared (0.09) and a standard alpha value (0.05) and power level (0.8) would be 42 participants with a critical F-value of 2.68 (actual results from observed data: $F=1.7$; $p=0.18$; $\eta_p^2=0.05$), which, when compared to the 32 participants tested across experiments six and seven, again support the notion that the interaction is potentially lacking the power required to achieve a significant interaction with a further five participants required per experimental group.

A remaining question is why no significant effect of offline rTMS was found in experiments three, four and five despite numerous changes in intensity, duration and frequency of the TMS pulses and previously reported success of using offline rTMS to disrupt speech perceptual abilities (Alba-Ferrara et al., 2012; Krieger-Redwood et al., 2013). This discrepancy could be due to a number of factors; initially it was hypothesised to be due to the overall lower intensity of the TMS pulses in experiment three compared to previous research (40 percent in experiment three compared to an average of 61 percent in the published literature), however the use of active motor threshold as a gauge of cortical responsiveness and subsequent higher average stimulation intensity in experiment four produced the same non-significant effects as experiment three. Whilst it is possible that the stimulation intensity in experiment four

was still too low to produce a significant effect, there was no noticeable movement towards significance in experiment four compared to experiment three which is more indicative that the intensity of the stimulation is not the sole determinant of the non-significant result.

The site of stimulation in the set of experiments presented here corresponds more to middle-anterior STS as opposed to the more posterior site of stimulation used by Krieger-Redwood et al. (2013), Alba-Ferrara et al. (2012) and others (Bestelmeyer et al., 2011; Grabski et al., 2013; Hirnstein et al., 2013; Meister et al., 2007). It is possible therefore that no effect was found in the first three experiments of this chapter due to stimulating the “wrong” region of cortex. However, this is unlikely as an effect of rTMS was found in experiment six using the same cortical location as the first three experiments. Conversely it is possible that the result from experiment six is a false positive and the null results of experiments three to five are more accurate and real. However, the sites of stimulation (left and right STS) were chosen as they were found to be the sites most consistently activated in a meta-analysis of functional imaging studies investigating the neurobiology of speech perception (Adank, 2012), it is unlikely therefore that this site is not involved in perception of speech and more likely that the result of experiment six is a real result and not a false positive (as supported by substantial evidence in favour of H_1 as indicated through Bayesian analyses).

A final possibility is the difference in the nature of the dependent variable used in the current set of experiments compared to previous experiments. The dependent variable of the SRT task is the average SNR level at which a reversal occurred, i.e., the point at which participants are able to or unable to repeat three or more of the presented key words correctly. Such a dependent variable is reflective of how accurately participants perceived (and then repeat) the heard sentence. Yet TMS is known to be more effective at disrupting response time measures more than accuracy measures (Alba-Ferrara et al., 2012; Hirnstein et al., 2013; Krieger-Redwood et al., 2013). This reflects the subtle effects of TMS where the cortical modulation is enough to delay participants in their judgments, yet not intense enough to entirely impair accurate decision making. Given the lack of a response time measure in the current set of experiments, it is possible that the offline protocols used in experiments three, four and five were delaying participants ability to perceive and accurately report speech in

noise but that the disruption was being missed due to the lack of a response time measure. The significant effect on accuracy of performance in experiment six is believed to be due to online TMS maximising cortical modulation within the superior temporal sulcus during presentation of the sentence without providing enough time for the intra- or interhemispheric compensation to moderate the behavioural effects of this modulation.

In addition to establishing the most effective TMS protocol for disrupting cortical processing of sentences in noise in comparison to the protocols used in experiments three to five, the results of experiment six also inform the ongoing debate in the neurobiology of language literature concerning the laterality of the speech perception network. The results show a TMS-induced impairment in speech perception after stimulation of both left *and* right temporal lobes and thus supports neurobiological models of speech perception that hypothesise bilateral processing in speech perception. These results have important ramifications for current and future neurobiological models of speech perception which should acknowledge and subsequently understand the important roles that both hemispheres play. Despite the equivalent level of disruption caused by the application of TMS it is not necessarily inferred that the processes being manipulated across the two hemispheres are equivalent. A symmetrical disruption does not in itself necessitate symmetrical functioning (Obleser et al., 2008; Scott et al., 2000) and several previous studies have argued in favour of hemispheric asymmetries in speech related auditory processing. In an fMRI study designed to specifically investigate the cortical mechanisms involved in the perception of speech in noise, Wong et al. (2008) found that speech embedded in noise resulted in increased activation in bilateral superior temporal gyrus. However, the pattern of activation differed between hemispheres. In the left STG activation continued to increase as the noise became more intense (from clear speech to an SNR of +20dB to an SNR of -5dB). However, in the right hemisphere activation increased from clear speech to the moderate SNR condition (+20dB) but did not increase any more as the noise became even more intense. Despite the selective nature of the right hemisphere change in activation Wong et al. (2008) found the degree of individual difference in the right hemisphere activation to be positively correlated with performance on a behavioural task in the most extreme listening condition

(participants with greater right hemisphere activation performed better on the behavioural task), with no correlation found between behavioural performance and left STG activation. When combined with the results of experiment six this suggests that speech perception is a bilateral process with both the left and right hemispheres performing important roles in the process, but the roles being performed by the hemispheres most likely differ.

Using magnetoencephalography (MEG) Shtyrov et al. (1998) found an increased magnetic mismatch negativity (MMNm) response in the right hemisphere when the speech signal was masked by noise with no change in the MMNm response occurring in the left hemisphere. The authors argue that this provides evidence for a redistribution in sound discrimination processing between the hemispheres as a result of the added difficulty produced by the addition of the background noise. Zatorre and Belin (2001) instead argue that the processing is not redistributed with changes in background noise but instead whilst bilateral posterior superior temporal areas respond to temporal variations and bilateral anterior superior temporal regions respond to spectral variations, the weighting of activations is such that the left always preferentially activates for temporal processing whilst the right is always weighted more towards spectral processing. In agreement with this hypothesis Obleser et al. (2008) observed minor differences in activation using fMRI when either the temporal or spectral details of speech were modified. On the whole spectral manipulations of the speech signal led to a small rightward lateralisation in processing whilst temporal manipulations revealed a small leftward lateralisation. When considering the specific activation of each hemisphere to both manipulations, the right hemisphere showed a preference for spectral details whilst the left hemisphere showed no real preference between either the spectral or the temporal manipulations. This is a pattern of activation that is reflected in many of the aforementioned functional imaging studies that find bilateral activation for intelligible speech. In these studies, activation in the left hemisphere is generally more selective to just intelligible speech and thus perhaps performing higher level linguistic processes whilst the right hemisphere is active for both intelligible and spectrally/temporally matched unintelligible speech. It is suggested that the increased general activation in the right hemisphere is indicative of mid-level acoustic-phonetic processes as opposed to the higher level processes of the

left (Evans, Kyong, Rosen, Golestani, Warren, McGettigan, Mourão-Miranda, et al., 2014; Friederici et al., 2010; Okada et al., 2010; Rosen et al., 2011; Scott et al., 2000). Obleser et al. (2008) argue that the results support the notion that the left and right temporal lobes both carry out critical processes for successful speech perception and contribute to our unified perception of speech in a different yet complimentary fashion. The results of experiment six showed that TMS can affect sentence processing in the bilateral temporal cortex and have shown the importance of each hemisphere in this task, based on this result future research should utilise TMS to focus on uncovering the specific role(s) that each hemisphere plays in the perception of speech.

Conclusions

In conclusion, the five experiments presented here provide an overview of the impact that different TMS design choices have on the outcome of an experiment. Across the five experiments a number of key design parameters were varied in an attempt to establish the most effective TMS protocol to modulate perception of sentences in noise. In varying the parameters, the presented experiments highlight a number of issues that must be considered when designing TMS experiments, i.e., the validity of using the threshold of the motor cortex as a proxy of whole brain responsiveness to TMS, increasing the intensity is not always the sole solution. Most critically of all these experiments highlight that perhaps the most important detail in any TMS experiment is the timing of the pulses. If the pulses are delivered at the wrong time point (be it online vs offline, or chronometric), researchers will either miss the cognitive function they are trying to modulate or will provide the brain with enough time to recover functioning through compensatory networks resulting in a masking of the modulation. Finally, the results of experiment six and seven show that despite potential confounding from the concurrent side effects of TMS, online rTMS can be used by researchers investigating the neurobiology of speech perception. Importantly, the online rTMS paradigm established to be most effective here will be used in the fourth chapter of this thesis to investigate the role of the ventral premotor cortex in adaptation to time-compressed speech.

Chapter Four

Investigating the Role of the Ventral Premotor Cortex in Adaptation to Time-Compressed Speech.

Introduction

The ventral Premotor cortex (PMv) is traditionally considered to be responsible for the planning, selection and initiation of movements (Iacoboni, 2008; Wise, 1985). However, a growing body of research also associates this region with other cognitive tasks, most notably, speech perception (Adank & Devlin, 2010; Meister et al., 2007; Sato, Tremblay, & Gracco, 2009; Schubotz & von Cramon, 2003). Initially the link between motor cortex and speech perception was suggested by the Motor Theory of Speech Perception (MTSP), which posited that we perceive speech through the internal simulation of heard sounds, i.e., our brain responds as if we were the one talking (Lieberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985). The theory of internal simulation was believed to be the case due to the issue of co-articulation where every word or sound is affected by the immediately adjacent words and sounds, for example, the /d/ in /di/ has a formant frequency that starts at a high frequency and rises, whilst the /d/ in /du/ starts with a low formant frequency and falls. The two /d/ phonemes thus have completely different acoustic profiles and yet in both cases the /d/ is perceived as the same sound (Lieberman et al., 1967). In order to maintain the perceptual invariance despite vast acoustic variation, Liberman et al. (1967) argued that an innate module existed that was separate from any of the general speech/motor regions and was devoted to the detection and processing of the intended gestures of the speaker, i.e., the movement of the articulators. In the example of the acoustically varying /d/ there is a commonality such that in both cases the tongue is always at the roof of the mouth. This led Liberman and Mattingly (1985) to propose that the basic units of speech perception are the invariant motor commands which are not only used to produce a phoneme/word but are also vital in perceiving the speech of others. Such invariant motor commands include factors such as ‘lip rounding’ and ‘jaw raising’ and Liberman and Mattingly (1985) argue that “*to perceive an utterance, then, is to perceive a specific pattern of intended*

gestures” (Lieberman & Mattingly, 1985, p.3) and thus acoustic features were not involved in the perception of speech.

In the intervening years, support for the most rigorous versions of MTSP have been falsified with evidence suggesting that not all speech sounds can be reduced to a single set of gestures but in some cases a many-to-one relationship exists (Schwartz, Basirat, Ménard, & Sato, 2012); categorical perception is not a trait unique to humans beings but also present in other animals such as Chinchillas (Kuhl & Miller, 1975); infants that are not yet able to produce speech sounds show signs of discrimination between different speech tokens (Eimas, Siqueland, Jusczyk, & Vigorito, 1971); and whilst damage to frontal and motoric regions produces significant production deficits, errors in perception are inconsistent and limited (Boatman, 2004; Stasenko, Garcea, & Mahon, 2013). All of which suggests that successful speech perception is not solely dependent on the processes of an innate module that perceives speech through the detection of invariant motor commands. Contrary to the motor theory, Scott, McGettigan, and Eisner (2009) propose that the motor cortex is not essential for speech perception, instead they ascribe this role to the classic posterior temporal regions. Instead they claim that the motor cortex is active during conversation as it entrains to the syllabic rhythm and speech rate of the current speaker. By keeping track of this information, the motor system helps in the specific act of turn taking. Specifically, the oscillatory entrainment of motor neurons ensures that when one speaker ends their turn, the interlocutor is ready to start talking. Scott et al. (2009) point to research that shows even when complete strangers talk over the telephone (that is, without any relevant visual input) over 85 percent of turn transitions occurs within a window of -750 to +750 milliseconds, representing what they term an “...astonishing level of coordination...” (p. 300) which, according to Scott et al. (2009) is guided by the motor cortex. The motor and premotor cortices therefore are active during speech perception in preparation for producing speech and not to assist with perception.

Subsequent research and the discovery of mirror neurons in the macaque homologue of the ventral premotor cortex (di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992; Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Kohler et al., 2002; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996), appear to suggest however, that the motor cortex plays a more extensive role than this theory advocates. Studies using

functional magnetic resonance imaging (fMRI) have shown that when subjects passively listen to monosyllables 73 ± 7 percent of voxels that appear active in superior ventral premotor regions (sPMv) are also active when participants are asked to reproduce those exact monosyllables, suggesting an overlap between perception and production in this region (Wilson, Saygin, Sereno, & Iacoboni, 2004). In addition, the activation in the sPMv region was reduced when participants heard non-speech sounds relative to speech stimuli. The selective activation of the premotor regions for speech perception suggests that the activation is stimulus-specific, suggesting that activation of premotor regions contributes to the analysis of speech stimuli. However, the precise role of sPMv in speech perception is unknown.

The Perception-for-Action-Control Theory (PACT; Schwartz et al. (2012) argues that the perceptual system is fundamentally linked to the motoric actions/gestures of speech as a result of both perception and production developing together during early childhood. Due to this intrinsic link between perception and production, the motor systems are provided with auditory templates that can be used to shape the internally stored gestural representations. However, unlike the traditional motor theorists, Schwartz et al. (2012) argue that the gestural representations are not purely motoric, but instead are shaped by perception and therefore refer to them as perceptuo-motor units. According to PACT, all speech perception is initially shaped by (auditory) perception. If the assistance of the motor cortex is required it is most likely to be under adverse listening conditions when the speech signal is distorted and the initial perceptual processes are insufficient (for example, when perceiving speech in noise or accented speech). In this context, Schwartz et al. (2012) suggest that the perceptuo-motor units stored within motor regions are activated to provide a better description of possible auditory templates to improve analysis of the auditory scene.

Evidence to support the theory that premotor activation and involvement in speech perception is context-dependent, has been observed in numerous functional imaging studies. Wilson and Iacoboni (2006) asked participants to listen to a series of phonemes that varied in both nativeness and producibility and found that activation in the sPMv increased for phonemes compared to rest and more specifically for non-native phonemes compared to native tokens. This result is predicted by PACT as perception of the nonnative phonemes would represent an example of more adverse

listening condition. However, activation in the sPMv did not vary as a function of producibility, a finding that was replicated by Tremblay and Small (2011). Yet activation in bilateral superior temporal and left superior-temporal-parietal regions did vary with producibility, i.e., as the phonemes became harder to produce, activation in these temporal regions increased. These results suggest that only bilateral superior temporal regions are crucial for mapping acoustic input onto phonetic codes. However, the fact that the sPMv areas showed differential activation for native and non-native phonemes suggests it is sensitive to phonemes that are (or are not) in the speakers repertoire, which in turn supports the notion that the PMv is relevant for speech perception. Wilson and Iacoboni (2006) hypothesise that the role of the premotor system is to generate top-down internal models of the incoming phonemes. When hearing clear, undistorted, speech sounds (e.g. native sounds), the motor cortex readily produces a template for the known phoneme. However, when faced with speech that has been distorted, it is possible that an internally stored template for the perceived phoneme may not exist. As a result, the premotor regions need to continuously search for and attempt to generate the closest possible internal model of the distorted speech and activation in this region increases. This theory is supported by Tremblay and Small (2011) who argue that if the role of the premotor cortex is simply to internally enact the same neural circuits involved in producing the speech tokens that are being perceived, then activation should be identical in production and perception. Whilst Tremblay and Small (2011) observed changes in activation dependent on the producibility of different consonant-vowel clusters during production, no modulation was observed during perception. As a result, Tremblay and Small (2011) agree that activation within premotor regions during speech perception is context- and task-dependent. They suggest that premotor activation occurs during speech perception as a consequence of spontaneous changes in the environment or task-requirements, e.g. increased background noise or introduction of a speaker with an unfamiliar accent that trigger a recalibration of the perceptual system to increase or decrease the amount of information from the premotor cortex. This conclusion suggests that activation in premotor regions is involved in speech perception but predominantly in a secondary, supportive role that is of more importance when the incoming signal becomes harder to perceive.

As stated in the first chapter of this thesis, activation in the ventral premotor cortex is observed during perception of noise-vocoded words (Hervais-Adelman et al., 2012) and CV syllables presented in noise (Osnes et al., 2011). Both studies found significantly greater PMv activation for the distorted but intelligible speech compared to clear speech, which supports the notion that PMv activation is context and/or task dependent. However, both Osnes et al. (2011) and Hervais-Adelman et al. (2012) argue that the speech related PMv activation represents a stimulus triggered process and that activation in the premotor regions automatically occurs in adverse speech listening conditions and is not an additional process that participants decide to perform due to the nature of the task.

The selective role of the ventral premotor cortex as a region that assists in the perception of degraded speech with limited assistance during perception of easy to comprehend speech was examined by Adank and Devlin (2010), who investigated the cortical regions involved in adaptation to time-compressed speech. At the start of the experiment, participants had no previous experience with time-compressed speech and behavioural results indicated that participants found this condition harder to comprehend, with slower response times and a higher number of errors relative to the clear speech baseline. Adank and Devlin (2010) found that the difficulty comprehending the time-compressed sentences co-occurred with increased activation in the bilateral pSTG and the left ventral premotor (IPMv) cortex. However, as participants adapted to the acoustic manipulation throughout the experiment, indicated by increasingly faster response times and a reduction in the number of errors made, the pattern of activation within these cortical regions also changed. During the first 16 sentences, representing the period of initial exposure, ventral premotor activation was significantly larger in response to time-compressed speech than clear speech. However, as performance on the task improved (suggestive of perceptual learning), the levels of activation within the IPMv cortex gradually declined and eventually returned to a level comparable to the clear speech condition (within 48 sentences). This finding supports the notion that the premotor cortex is involved in the perception and adaptation to distorted speech stimuli and shows for the first time how ongoing activation within the PMv region may be modulated by task demands.

As noted previously however, the main limitation of functional imaging research is that all results are based on correlational designs by linking changes in blood oxygen levels in a cortical region with task performance. TMS, on the other hand, enables conclusions of causation, with disruption of behavioural performance after the application of TMS indicative of the functional relevance of the region being stimulated. Thus far, research investigating the role of the ventral premotor cortex in speech perception using TMS has been limited but in general supports the claim that the ventral premotor cortex contributes to processes of successful speech perception. Meister et al. (2007) found a significant impairment in the discrimination of voiceless stop consonants presented in the presence of background noise following the application of rTMS to the left ventral premotor region, with no effect of rTMS found for a task requiring the discrimination of tones. Additionally, Krieger-Redwood et al. (2013) asked participants to make phonological (was the final phoneme /p/ or /t/?) or semantic (is the item man-made or natural?) judgments following rTMS to either the left pSTG, lPMv or the occipital pole (control site). Krieger-Redwood et al. (2013) found that application of rTMS to pSTG impaired both phonological and semantic discrimination, however rTMS applied to the ventral premotor cortex only impaired response times on the phonological task. As a result, the authors conclude that the left PMv may only be crucial for tasks that require phonological segmentation and explicit phonemic awareness but activation in the PMv is not crucial for successful comprehension. This is in agreement with results from Sato et al. (2009), who investigated the role that the PMv plays in the perception of clear speech tokens. Participants were asked to complete a phoneme identification task (is the syllable /p/ or /b/?); a syllable identification task (/put/ vs /but/, same syllable or not?) or a phoneme discrimination task (/put/ vs /bon/, same initial phoneme?) after receiving offline rTMS, Sato et al. (2009) found a significant effect only on the task requiring phonemic discrimination. On the basis that the stimuli in the phoneme identification and syllable discrimination tasks were equal except for the first phoneme, Sato et al. (2009) argue that these tasks can largely be completed solely using acoustic analyses in auditory brain regions (motor regions are not required). Whilst the task in the phoneme discrimination condition was to discriminate the initial phoneme, in this condition phonemes in addition to the initial phoneme could differ between speech tokens, e.g., in /put/ vs /bon/ all phonemes, not just the initial ones, differ. The authors

argue therefore that, in addition to acoustic analyses, successful completion of this task requires explicit segmentation of the speech signal into its constituent parts to establish what the initial phoneme was. The fact that this condition was the only task to be impaired by the application of rTMS to the ventral premotor regions, led Sato et al. (2009) to conclude, in agreement with Krieger-Redwood et al. (2013), that the role of the PMv cortex is to assist in the segmentation of the speech signal into its constituent phonemes, a process it can perform dependent on task demands, in both clear and distorted conditions.

In summary, PMv has traditionally been associated with the planning, selection and initiation of bodily movements however a growing body of research has started to associate this region with a role in speech perception. Whilst initial theories suggested that PMv may play a critical role in successful perception of speech, subsequent research suggests that PMv plays a secondary, non-critical role by assisting on tasks requiring explicit phoneme segmentation. Furthermore, it is hypothesised that this role is context-dependent with greater importance placed on the processes of this region during adverse listening conditions. This is supported by the work of Adank and Devlin (2010) who found a modulation of PMv activation as participants adapted to time-compressed speech. The authors conclude that the modulated PMv activation is a neural signature of perceptual adaptation to distorted speech. However, due to the correlational nature of functional imaging data it is unknown whether the activation in PMv is of critical, secondary or no importance to perceptual adaptation to distorted speech.

The aim of the present experiment is therefore to determine the extent to which the left PMv cortex is involved specifically in adaptation to distorted speech stimuli. To this end, using a between-subjects design, four groups of participants were asked to perform a computerised version of the SCOLP (the same test used by Adank & Devlin, 2010) test across two speech conditions per group. All four groups of participants completed the SCOLP test for 60 sentences in a clear, uncompressed form and for 60 sentences that had been time-compressed to 40 percent of their original length. As adaptation to time-compressed speech can only occur once, four separate groups of participants were used in order to be able to apply TMS to multiple cortical sites in the left hemisphere and measure the effect on perceptual adaptation to the

distorted stimuli. The four groups of participants therefore correspond to the four different sites of TMS stimulation used in the experiment: left superior temporal sulcus (STS); left ventral premotor cortex (PMv); left lateral occipital cortex (LOC) and a final group/condition where no TMS was applied directly to the cerebral cortex (no TMS). As discussed previously, TMS is known to be more effective at disrupting response times more than accuracy scores and therefore the SCOLP test, as used in experiments one and two of this thesis, was preferred in the current experiment to the SRT task used in experiments three to seven. Additionally, time-compressed speech was used as the sole speech distortion as it not only represents a type of acoustic distortion that is predicted to result in recalibration of the perceptual systems and engagement of the premotor cortices but additionally is a type of distortion that is associated with engagement of phonological segmentation processes (Sebastián-Gallés et al., 2000). Furthermore, evidence of adaptation, in terms of reductions in response times in experiments one and two, was greatest for time-compressed speech and occurred in the shortest period of time. This is important as it limits the number of trials required to observe adaptation and therefore restricts the number of TMS pulses that are required to be administered to participants, thus making the study safer and more comfortable.

Overall it was predicted that no effect of TMS would be observed in the no TMS and LOC groups for either the clear or time-compressed speech conditions. Given the subtle effects of rTMS and the ease of the clear speech condition, no effect of rTMS was predicted during stimulation of left STS or PMv groups either, however if an effect was to occur it would only be expected in the left STS group. In comparison, application of online rTMS was expected to induce overall significantly slower response times for the time-compressed condition throughout stimulation for the left STS group. Finally, modulation of performance induced through stimulation of the PMv site was only expected to occur whilst participants were adapting to the distorted speech signal (within the first 30 sentences of this condition). Once adaptation had occurred, no effect of rTMS to PMv was expected (i.e., no effect was expected on the final 30 sentences of this condition).

Experiment Eight

Methods

Participants

Forty participants (mean age 21yrs 3mths \pm 2.23; range 18-26; 29 females) were recruited for this experiment. All participants were native British English speakers, had normal or corrected to normal vision and were right-handed as assessed through dominant writing hand. No participants reported a history of speech, language, neurological or psychiatric disorder. All participants were assessed to have normal hearing, i.e., pure tone threshold of 20dB HL or better at octave frequencies between 250 and 8000Hz in both ears (British Society of Audiology, 2011). No participants presented with any contraindications for TMS, all gave informed consent and were paid for their participation.

Table 22 - Summary Statistics of the four different groups in Experiment Eight.

Group	Age	Range	F:M	PTA (L)	PTA (R)	SRT
STS	22.9 \pm 1.66	21-26	7:3	1.25 \pm 3.09	1.50 \pm 4.11	-3.89 \pm 0.88
PMv	20.7 \pm 1.88	18-23	8:2	6.11 \pm 5.23	6.94 \pm 4.87	-3.13 \pm 1.06
LOC	21.7 \pm 2.54	18-16	9:1	1.58 \pm 4.33	2.58 \pm 2.37	-3.31 \pm 1.06
No TMS	20.1 \pm 1.96	18-24	5:5	3.50 \pm 6.05	3.16 \pm 5.32	-3.17 \pm 1.33

Notes: F:M indicates the Female to Male ratio participant gender per group. PTA represents the hearing threshold in decibels averaged across all measured frequencies for each ear separately (octaves from 250-8000Hz). SRT represents the Speech Reception Threshold signal to noise level in decibels.

Procedure

To assess the role of the left ventral premotor cortex in adaptation to distorted speech, participants were asked to complete a computerised version of the SCOLP test, the same speeded sentence verification task as used in experiments one and two. Participants listened to simple sentences in two conditions: (1) clear, uncompressed speech and (2) time-compressed speech and had to decide whether the sentence they heard was true or false, indicating their response by pressing either the left (true) or right (false) key of a standard PC keyboard. All sentences were clearly true ('*Admirals are people*') or false ('*Admirals have fins*'). Accuracy and response times were recorded per trial with adaptation to each condition adjudged via improvements in

speed and accuracy of sentence verification. As adaptation is only possible once, this experiment consisted of four groups of 10 participants. Each group of participants received online rTMS to a single cortical region with the site of stimulation changing between groups. The final analysis compares performance in the clear and time-compressed conditions between the four groups / TMS conditions (see Table 22 for summary statistics of the four groups).

Stimuli

The stimuli were recordings of 120 SCOLP sentences, 60 true and 60 false (see Appendix N). Half of the sentences were presented in the clear speech condition and

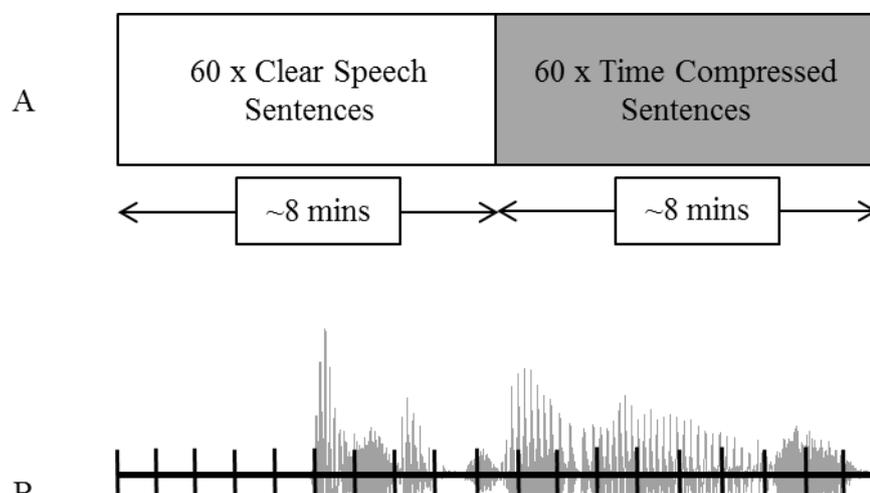


Figure 16 - A - Illustration of the order of events and associated timings for Experiment Eight. B - Illustration of a single trial and associated timings of TMS pulses for Experiment Eight. Vertical dashes represent each occurrence of a TMS pulse.

half in the time-compressed condition. Sentences varied from three to seven words (mean length 3.86 ± 0.87) with an average of 6.35 syllables per sentence (range 3-16) and an average length of 1.11 seconds (SD 0.16; range 0.73-1.41 seconds). All sentences were recorded by the same male speaker of standard British English. The speaker corresponded to speaker four in experiments one and two. Participants were overall slower to respond to this speaker in experiments one and two however response times in the time-compressed condition reduced (i.e., became quicker) from the first to second, third and fourth block more for this speaker than for the other three speakers. This suggests that adaptation to time-compressed sentences from this speaker was

possible, and therefore TMS induced modulation of adaptation would also be possible. All sentences were saved to separate files with the beginning and end trimmed to zero crossings as closely as possible to the onset/offset of the initial/final speech sounds; resampled to 22050 Hz; peak normalized to 99% of maximum amplitude and scaled to 70dB SPL using Praat (Boersma & Weenink, 2011). Stimulus presentation was performed using a custom-made MATLAB 2014a program (The MathWorks Inc., Natick, MA, 2000) and Etymotic ER-1 insert earphones, with all stimuli delivered at a comfortable listening level (preset at 74dB SPL but where necessary this was adjusted to fit individual participant preference).

Time-compressed sentences were shortened to 40% of their original length, resulting in an average syllable rate of 14.21 syllables per second (clear speech: 5.68 sy/sec). Time compression was implemented using the same PSOLA algorithm in Praat as used in experiments one and two (Charpentier & Stella, 1986). In contrast the clear, uncompressed, sentences were presented without any manipulation (beyond the zero trimmings, peak normalization etc. outlined above).

Stimuli presentation was blocked by condition with all 60 sentences from one condition being presented together, followed by the 60 sentences from the other condition. Half of the participants heard the uncompressed stimuli first followed by the time-compressed stimuli, whilst the other half of participants heard the time-compressed sentences first followed by the 60 uncompressed sentences.

Transcranial Magnetic Stimulation

Stimulation was performed using a Magstim Rapid² and a 70mm figure-of-eight coil (Magstim, Whitland, UK). Pulses were delivered online (i.e., at the same time as participants were listening to the sentences) at a rate of 10Hz. Stimulation intensity was set at 40% of the maximum pulse strength and held constant across all participants. Before the experiment began each participant was randomly allocated to a TMS condition. Upon arrival (and after safety screening) all participants received two trains of pulses on the site of stimulation that they were randomly allocated to ensure they were comfortable with the stimulation parameters. If participants found the site of stimulation particularly uncomfortable, they were reallocated to one of the other TMS

groups. During this demonstration, all participants used an earplug (3M E.A.R., 36dB attenuation) in the ear ipsilateral to the site of stimulation to attenuate the sound of the coil discharge and avoid damage to the ear (Counter et al., 1991). During the main experiment, magnetically shielded ER1 Etymotic earphones were used bilaterally to both deliver the auditory stimuli and attenuate the sound of coil discharge.

The number of TMS pulses delivered in each condition were adjusted to reflect the significantly different length of stimulus duration between the clear and the time-compressed sentences. In the clear condition, twenty pulses were delivered starting 500 milliseconds before sentence onset and running until sentence offset (15 pulses during sentence presentation – one every 100 milliseconds). Whilst in the time-compressed condition, ten pulses were delivered with stimulation again starting 500 milliseconds before sentence onset and continuing until the end of each sentence (5 pulses during sentence presentation – one every 100 milliseconds – see Figure 16 above). Whilst this difference in number of pulses between conditions does introduce a potential confound into the research it is a confound that is expected to be more likely to produce a false-rejection (type-two statistical error) of the experimental hypothesis rather than a false-positive (type one statistical error) acceptance of the experimental hypothesis. This is because the condition of most experimental interest is the time-compressed condition which is receiving the fewest number of pulses therefore making it harder to find an effect of TMS in this condition compared to the clear speech, baseline condition where twice the number of pulses are being delivered.

Site of Stimulation

In place of a participant specific structural scan the MNI-152 brain was used to guide placement of the TMS coil. In conjunction with Brainsight 2.3.5 the MNI-152 brain was adapted based on a minimum of ten separate estimations of the front-, back-, top, left- and rightmost points on each participants head with the MNI brain adapted to meet the measured dimensions. The same coordinates for the left STS group as experiments three to seven were used in the current experiment ($x = -60$, $y = -12$, $z = -6$) corresponding to a middle to anterior portion of the left STS. The coordinates for the left PMv group were taken from Adank and Devlin (2010) and correspond to the

coordinate of peak activation for the observed adaptation related changes in the left PMv in their experiment ($x = -50$, $y = +14$, $z = +12$). The LOC condition was included as a TMS control condition, this site was chosen specifically as application of repetitive pulses of TMS to the lateral occipital region induces a range of twitches in the head and neck muscles that can be distracting and sometimes uncomfortable. Additionally, to stimulate this region requires placement of the coil just behind the left ear and therefore the sound of coil discharge is equivalent to the STS (coil just above the left ear) and PMv (coil just in front of left ear) conditions. As a result, this location was considered as a fair comparison to control for the potentially confounding side effects of TMS. Coordinates for the LOC were taken from Duncan, Pattamadilok, Knierim, and Devlin (2009) and correspond to $x = -43$, $y = -77$, $z = -8$. The fourth group in the experiment included the 10 participants who received no TMS and this group served as a baseline against which the other three groups were compared. Whilst participants in this group did not receive full strength stimulation, the coil was held on the participant's head and pulses were discharged from the TMS machine at the same rate and intensity as in the other three groups. However, to prevent full stimulation the coil was rotated 90 degrees, with the side of the wing pressed against the participant's head. This was done to provide another condition against which the auditory click of coil discharge could be controlled for, as well as controlling for potential behavioural effects induced in participants as a result of having an unfamiliar scientific instrument held against their heads whilst completing the task.

Data Analysis

Half of the participants in the experiment were exposed to the clear sentences first followed by the time-compressed sentences and vice versa. Therefore, two, four-way, mixed repeated measures analysis of variances (ANOVA) were conducted to investigate the effect of the application of TMS to each of four different cortical sites on participants ability to accurately identify a clear or time-compressed sentence as true or false. TMS group (STS vs PMv vs LOC vs no TMS) and Order (Clear sentences first vs Time-compressed sentences first) were included as between-subject factors with Condition (Clear vs Time-compressed) and Block (the 60 sentences per condition

were post-hoc divided into 12 blocks of five sentences to establish adaptation) entered as within-subject factors. One ANOVA was conducted for the accuracy data and one for the response time data, only correct response trials were analysed in the response time ANOVA. All follow up analyses are subject to family-wise error rate corrections, specifically the Bonferroni correction.

Results

Accuracy

The results of the ANOVA revealed a significant two-way interaction between speech condition and block $F(11, 352) = 2.03, p = 0.02, \eta^2 = 0.05, B_{(1,0)} = 1.03$. Whilst no significant differences survived multiple comparison corrections ($\alpha = 0.05 / 11 = 0.004$), the results of paired samples t -tests suggest a delayed rate of improvement for the time-compressed speech condition compared to the clear speech condition. In the clear speech condition a significant change in accuracy was observed between the first two blocks $t(39) = -2.05, p = 0.047, CI = [-10.91, -0.08]$, Cohen's $d = -0.325, B_{(1,0)} = 1.12$, indicating significant improvement in the first five to ten sentences of the condition, likely reflecting procedural learning, i.e., familiarisation with the task and testing environment. In contrast, in the time-compressed condition a significant change in accuracy was observed between the second and third blocks $t(39) = -2.76, p = 0.009, CI = [-20.79, -3.21]$, Cohen's $d = -0.43, B_{(1,0)} = 4.58$, indicating that adaptation took slightly longer in this condition, with the most significant improvement in accuracy occurring after 10 to 15 sentences worth of exposure. This is in agreement with previous research (see summary of time-compressed speech research in Chapter Two) where adaptation to time-compressed speech has been shown to predominantly occur within the first 20 sentences of exposure.

In addition, a significant interaction between speech condition and order of presentation was found $F(1,32) = 13.001, p = 0.001, \eta^2 = 0.07, B_{(1,0)} = 1.07$. Independent samples t -Tests established that performance in the clear speech condition was unaffected by the order of presentation ($t(38) = -0.42, p = 0.67, CI = [-2.38, 1.55]$, Cohen's $d = -0.13, B_{(1,0)} = 0.33$). However, the order of presentation had a significant impact on perception of the time-compressed sentences ($t(38) = 3.21, p = 0.003, CI =$

[3.42, 15.07], Cohen's $d = 1.01$, $B_{(1,0)} = 14.105$) with participants performing significantly better when the time-compressed sentences ($M = 84$, $SEM = 1.74$) were presented after the clear speech condition as opposed to when the time-compressed sentences were presented first ($M = 74.75$, $SEM = 10.22$).

Finally, the ANOVA revealed a significant main effect of speech condition $F(1,32) = 132.27$, $p < 0.001$, $\eta^2 = 0.72$, $B_{(1,0)} > 1000$, with participants being more accurate in the clear speech condition ($M = 94.79$, $SE = 0.48$) than in the time-compressed condition ($M = 79.35$, $SE = 1.56$). A significant main effect of block was also found $F(11, 352) = 3.2$, $p < 0.001$, $\eta^2 = 0.07$, $B_{(1,0)} = 0.709$, this effect was driven by the significant improvement in accuracy from the first five sentences ($M = 80.75$, $SE = 2.13$) to the eleventh ($M = 90.25$, $SE = 1.57$) and twelfth blocks ($M = 90$, $SEM = 1.38$). As well as a significant main effect of order $F(1,32) = 6.902$, $p = 0.01$, $\eta^2 = 0.15$, $B_{(1,0)} = 1.105$, indicating that participants who were exposed to the clear sentences first ($M = 89.29$, $SEM = 1.18$) performed significantly better overall compared to the participants who were exposed to the time-compressed sentences first ($M = 84.87$, $SEM = 1.18$). No other main effects or interactions were found to be significant (all p 's > 0.25), i.e., overall no effects of TMS were found.

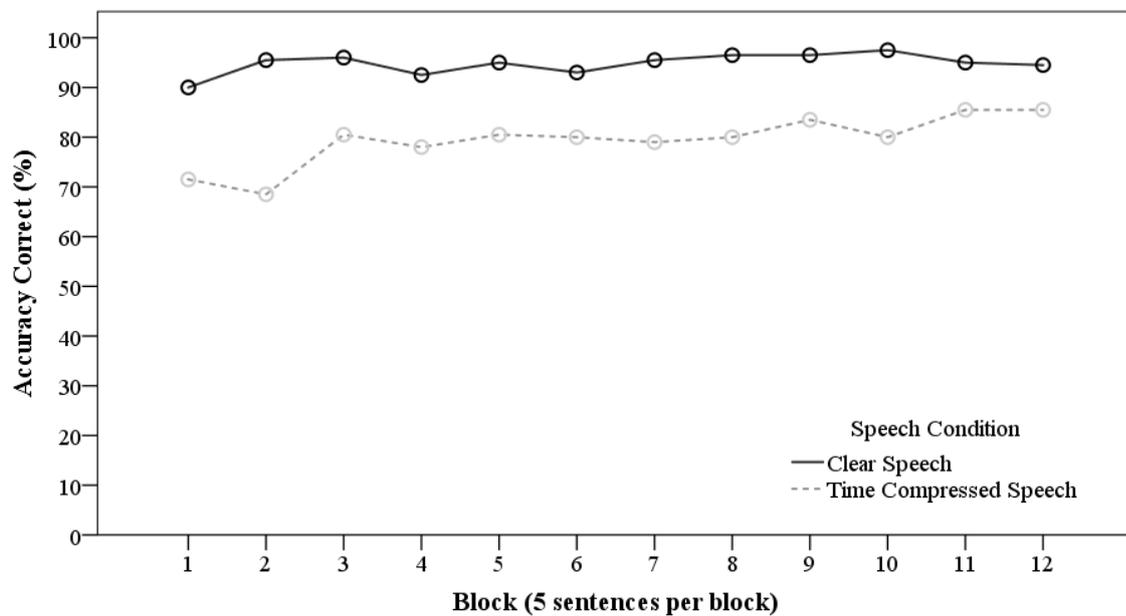
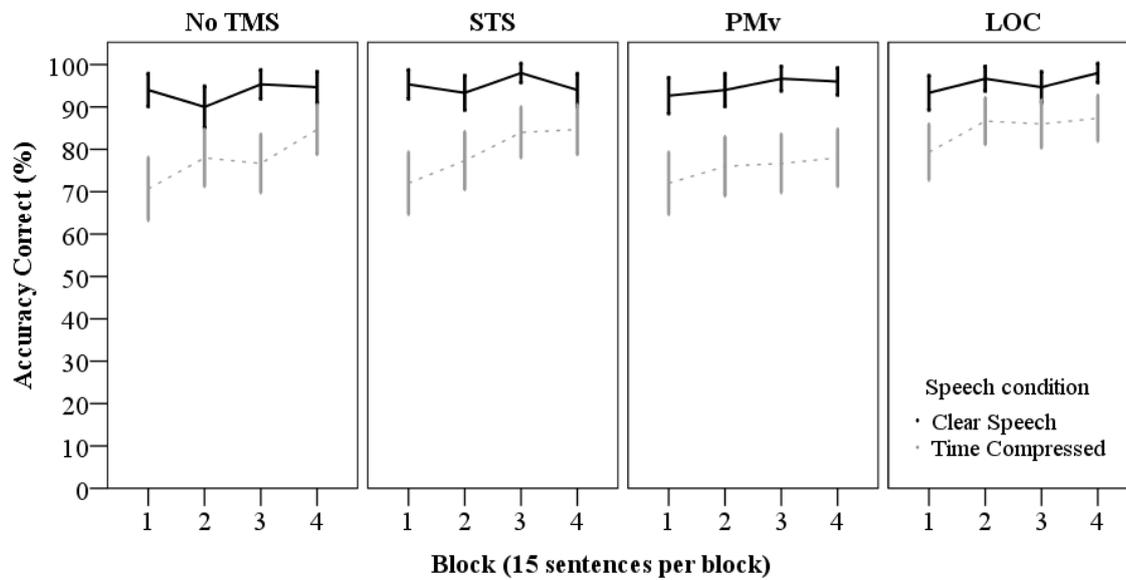


Figure 17 - A - Line graphs displaying Accuracy of performance for Experiment Eight across TMS group, Condition and Block Number. B - Line graphs displaying Accuracy of performance for Experiment Eight across Condition and Block Number.

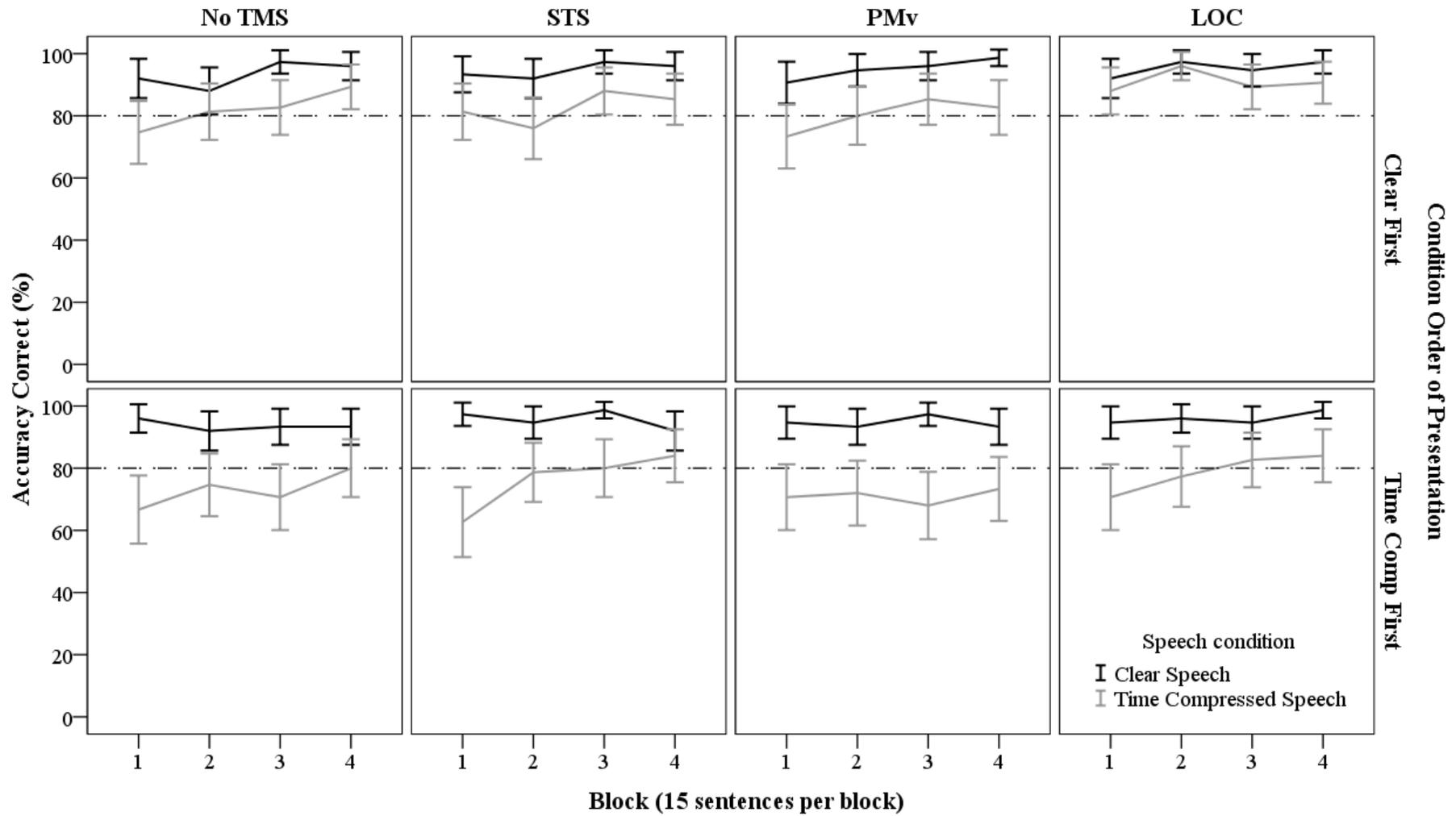


Figure 18 - Line graphs displaying Accuracy of performance for Experiment Eight across TMS group, Condition, Block Number and Order of Condition.

Response Times

For the response time data a significant two-way interaction between speech condition and order of presentation was also found $F(1,31) = 6.39, p = 0.017, \eta^2 = 0.02, B_{(1,0)} = 1.02$. Independent samples *t*-Tests established that performance in the clear speech condition was again unaffected by the order of presentation ($t(38) = -1.34, p = 0.18, CI = [-184.7, 37.39]$, Cohen's $d = -0.42, B_{(1,0)} = 0.62$). However, the order of presentation had a significant impact on perception of the time-compressed sentences ($t(38) = -3.33, p = 0.002, CI = [-394.4, -96.55]$, Cohen's $d = 1.05, B_{(1,0)} = 18.43$) with participants performing significantly better when the time-compressed sentences ($M = 849.43\text{ms}, SEM = 52.77$) were presented after the clear speech condition as opposed to when the time-compressed sentences were presented first ($M = 1094.88\text{ms}, SEM = 51.23$).

Additionally, the ANOVA revealed a significant main effect of speech condition $F(1,31) = 231.79, p < 0.001, \eta^2 = 0.84, B_{(1,0)} > 1000$, with participants quicker to respond accurately in the clear speech condition ($M = 463.07\text{ms}, SE = 28.69$) compared to the time-compressed condition ($M = 979.03\text{ms}, SE = 42.62$). Additionally, a significant effect of block was found $F(7.11, 220.64) = 5.42, p < 0.001, \eta^2 = 0.12, B_{(1,0)} = 0.018$, this was due to a significant reduction in response times from the first block ($M = 842.34, SE = 41.82$) to the second block ($M = 712.25, SE = 34.77$) before a relative asymptote. Finally, a significant main effect of order of condition presentation was found $F(1,31) = 6.09, p = 0.01, \eta^2 = 0.15, B_{(1,0)} = 3.08$, indicating that participants who were exposed to the clear sentences first ($M = 644.67\text{ms}, SEM = 43.58$) performed significantly better overall compared to the participants who were exposed to the time-compressed sentences first ($M = 794.54, SEM = 42.28$). No other main effects or interactions were found to be significant (all p 's > 0.25) again indicative of no effect of TMS at any target site or in either condition.

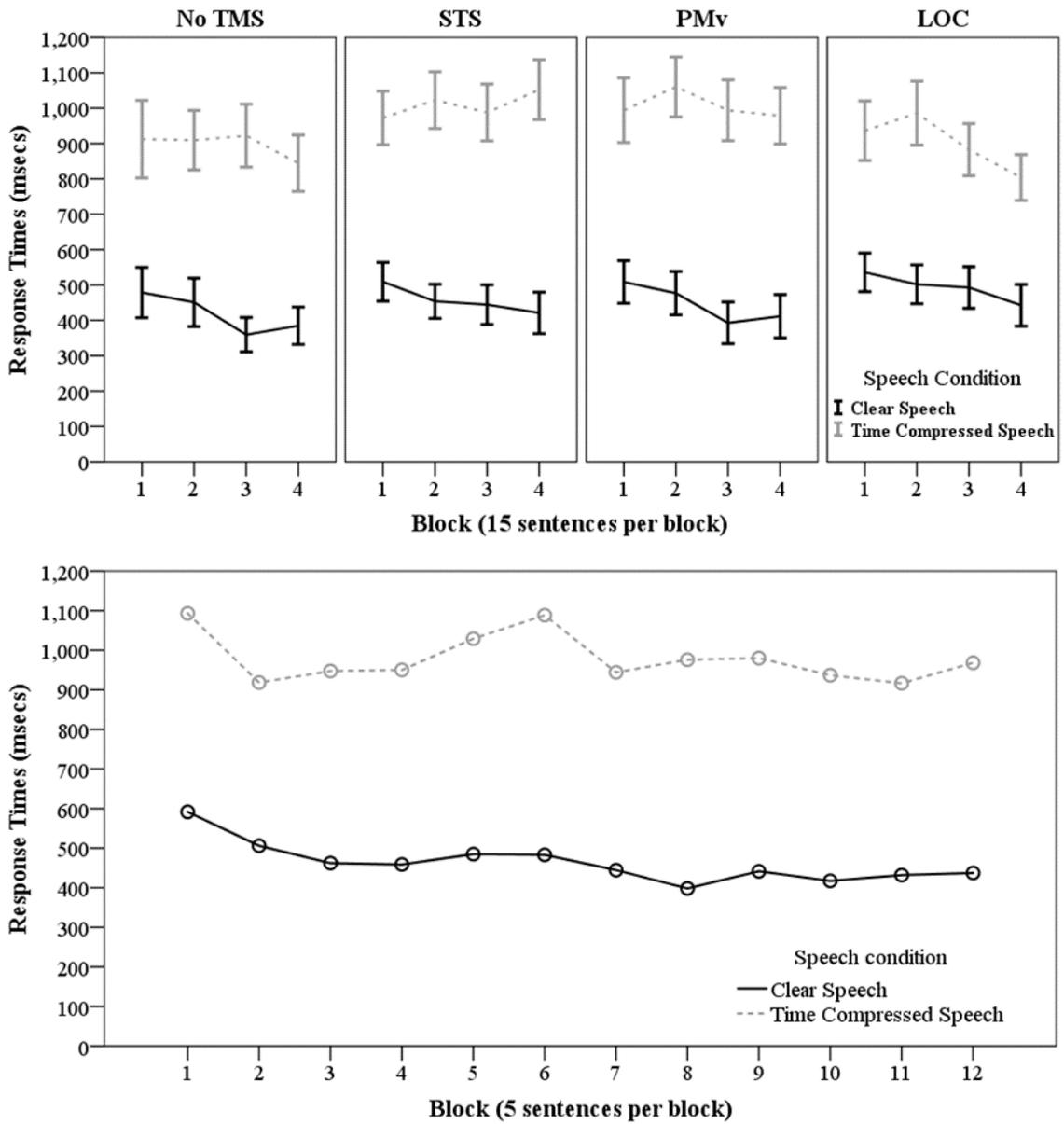


Figure 19 - A - Line graphs displaying Response Times performance for Experiment Eight across TMS group, Condition and Block Number. B - Line graphs displaying Response Time performance for Experiment Eight across Condition and Block Number.

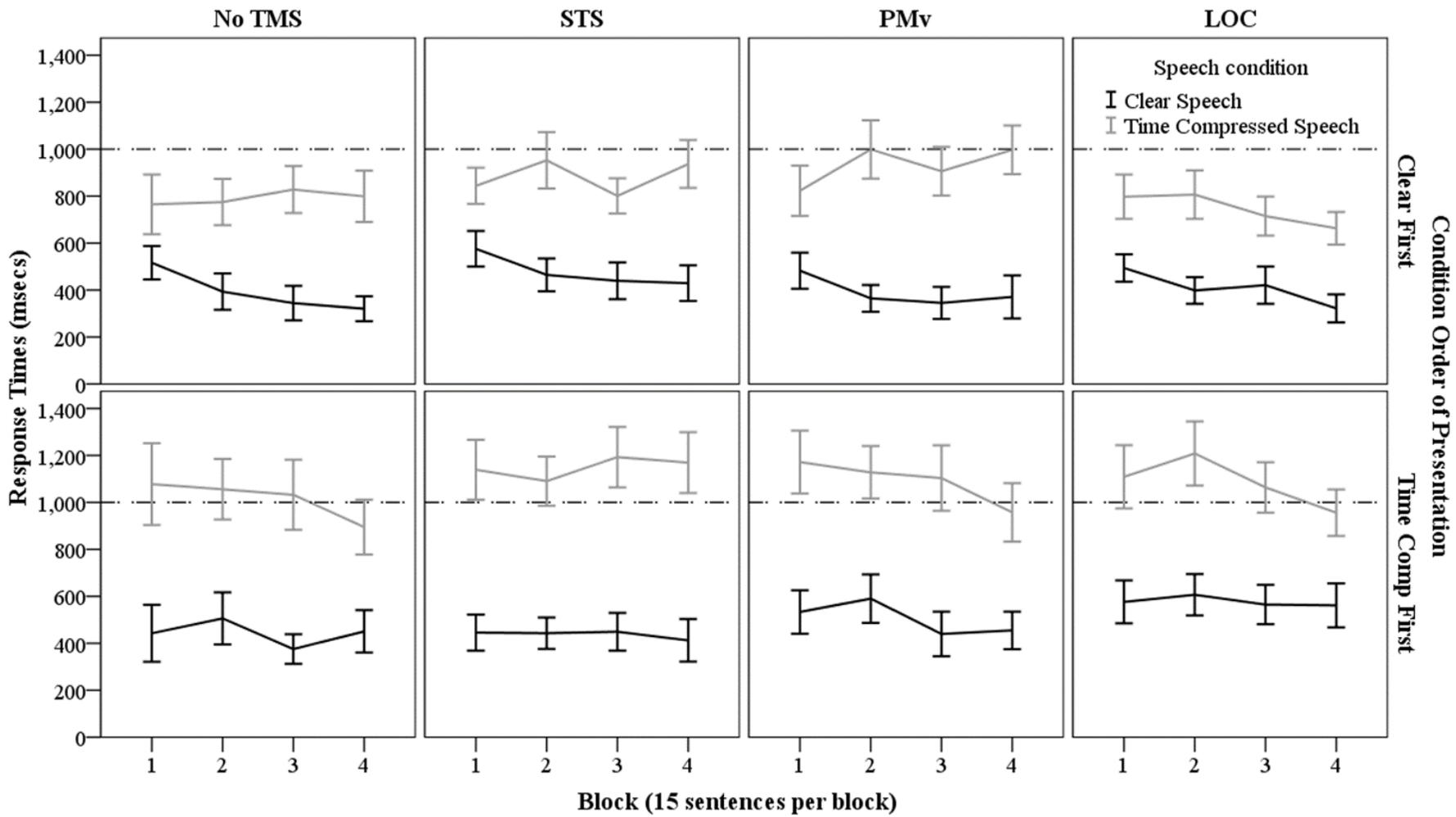


Figure 20 - Line graphs displaying Response Times performance for Experiment Eight across TMS group, Condition, Block Number and Order of Condition.

Table 23 - Means and Standard Deviations of the Accuracy and Response Time data for the two speech condition across blocks of 15 sentences.

Accuracy : Clear Condition					
	Block 1	Block 2	Block 3	Block 4	Overall
No TMS	94 (8.57)	90 (7.85)	95.33 (6.32)	94.66 (6.88)	93.5 (3.8)
STS	95.33 (8.91)	93.33 (6.28)	98 (3.22)	94 (7.98)	95.16 (3.08)
PMv	92.66 (3.78)	94 (4.91)	96.66 (3.51)	96 (6.44)	94.83 (2.65)
LOC	93.33 (5.44)	96.66 (3.51)	94.66 (6.88)	98 (4.49)	95.66 (2.5)

Accuracy: Time-Compressed Condition					
	Block 1	Block 2	Block 3	Block 4	Overall
No TMS	70.66 (14.12)	78 (9.92)	76.67 (19.18)	84.66 (14.41)	77.5 (8.68)
STS	72 (13.62)	77.33 (14.47)	84 (10.97)	84.66 (7.06)	79.5 (7.81)
PMv	72 (17.99)	76 (13.03)	76.66 (16.1)	78 (11.77)	75.66 (11.89)
LOC	79.33 (17.34)	86.67 (10.88)	86 (14.21)	87.33 (9.66)	84.83 (10.67)

Response Times: Clear Condition					
	Block 1	Block 2	Block 3	Block 4	Overall
No TMS	478.77 (224.64)	450.11 (219.62)	363.47 (172.85)	383.88 (202.24)	419.06 (186.65)
STS	523.28 (209.25)	451.96 (159.62)	445.55 (164.43)	425.27 (172.19)	461.52 (143.14)
PMv	512.38 (203.17)	475.14 (241.69)	388.35 (192.51)	418.29 (223.75)	448.54 (190.95)
LOC	534.68 (174.76)	500.43 (223.25)	497.83 (216.39)	443.51 (229.22)	494.12 (198.04)

Response Times: Time-Compressed Condition					
	Block 1	Block 2	Block 3	Block 4	Overall
No TMS	956.1 (431.55)	922.9 (273.54)	929.44 (321.17)	859.29 (231.5)	916.93 (295.9)
STS	1017.37 (323.36)	1029.31 (248.39)	995.33 (337.34)	1049.36 (303.07)	1022.84 (282.15)
PMv	1038.2 (291.95)	1064.45 (238.03)	1017.64 (252.73)	998.27 (235.83)	1029.64 (196.66)
LOC	950.68 (280.47)	1012.26 (391.85)	900.82 (275.37)	822.16 (218.78)	921.48 (279.58)

Notes: Numbers outside of parenthesis represent the mean values, numbers inside the parenthesis represent the standard deviation.

Discussion

The present study aimed to determine the extent to which the left ventral premotor cortex is involved in adaptation to time-compressed speech, through the application of rTMS. Research thus far has associated the ventral premotor region with a secondary and task dependent role in speech perception. Specifically, it is believed that the PMv cortex is strongly activated by adverse listening conditions and assists the speech perceptual systems through additional phonemic segmentation processes. Therefore, based on the results of Adank and Devlin (2010) an effect of rTMS (realised as overall slower response times) was expected during initial exposure and adaptation to time-compressed sentences (within the first 30 sentences) when targeted at the left PMv cortex. To this end, four groups of participants were asked to perform a SCOLP test for 60 clear, uncompressed sentences and 60 time-compressed sentences. Each group received online rTMS, however, groups differed in the cortical site targeted with participants receiving rTMS to either the left superior temporal gyrus, left ventral premotor cortex, left lateral occipital cortex (control site) or a no TMS baseline control group. Despite the prediction that TMS to left PMv would disrupt adaptation to time-compressed speech, overall, the results show no significant TMS group effect or interaction suggesting that the application of online rTMS did not impair perception of or adaptation to either of the speech conditions at any of the cortical sites of interest.

It is possible that the role of PMv in speech perception especially in adverse listening conditions is not of significant enough importance that modulation of these processes through application of rTMS, results in behavioural effects. Alternatively it is also possible that no effect of rTMS application to this region was observed because of its lack of functional relevance to speech perception (Scott et al., 2009). However, as outlined previously, performance on numerous speech perception tasks' has been impaired by the application of rTMS to the left PMv (Krieger-Redwood et al., 2013; Meister et al., 2007; Sato et al., 2009) suggesting that this region is involved in speech perception processes and TMS is capable of disrupting these processes, thus making suggestions of functional irrelevance hard to reconcile. Yet each of the previous studies have used a task that tested perception for individual words or phonemes and in the case of Sato et al. (2009) and Krieger-Redwood et al. (2013) relied specifically upon phonemic segmentation processes. It is possible therefore that the task used in

the present experiment did not place enough stress on the speech perceptual system to require the PMv to engage in supportive phonemic segmentation processes. However, adaptation to time-compressed speech is strongly associated with perceptual adjustments and shifts in attention at the level of phonological representations as supported by the generalisation of learning across different languages and speakers (Pallier et al., 1998; Sebastián-Gallés et al., 2000). Furthermore, as discussed in Chapter Two it is predicted that overall adaptation rates in this condition are associated with higher overall statistical learning abilities of participants (Neger et al., 2014; Palmer & Mattys, 2016). It is likely therefore that participants would have been engaged in attempting to identify regularities in the distorted speech signal during adaptation with specific focus on phonological segmentation, thus making the lack of a significant rTMS effect unexpected.

In addition to the lack of a significant effect of rTMS to the PMv site, this study found no significant effect of stimulating the left STS on participants' accuracy or speed of response to either the clear speech or time-compressed condition. Whilst no effect of stimulation in the clear condition was hypothesised due to the ease of the task and the subtle effects of TMS modulation, stimulation of the left STS was expected to significantly impair response times in the harder, time-compressed condition. It is possible that the lack of a significant effect at this target cortical site (MNI coordinates: -60, -12, -6) was due to stimulation of a part of the left STS that was not functionally relevant for speech perception or more specifically perception of speech in adverse listening conditions. The site of stimulation in the current experiment is in a middle to anterior portion of the left STS, yet, as outlined in Chapter One of this thesis, increased activation related to decreasing intelligibility is more commonly associated with the posterior STS, especially the planum temporale (Adank et al., 2012; Griffiths & Warren, 2002; Poldrack et al., 2001; Scott et al., 2006; Wong et al., 2008). Indeed, one of the regions showing adaptation related changes in the Adank and Devlin (2010) functional imaging investigation into the neural changes associated with adaptation to time-compressed speech was the left posterior STS (peak MNI coordinates: -54, -52, 2). It is possible therefore that no effect of TMS was observed in the present study due to the functional irrelevance of the site being stimulated. However, as noted in Chapter Three the site of stimulation was chosen because of its consistent activation across a range of functional imaging studies investigating the neurobiology of speech

perception (Adank, 2012). In addition, whilst the site being stimulated was more anterior than the site identified by Adank and Devlin (2010) as showing adaptation related changes, the STS region stimulated in the present experiment was close to a peak coordinate for an STS region found to be more active when participants heard speech relative to baseline (MNI coordinates: -58, -16, -6) and a site showing greater activation for the time-compressed speech relative to normal uncompressed sentences (MNI coordinates: -60, -14, 0; Adank & Devlin, 2010). Therefore, whilst it is possible that the STS region being stimulated in the current study may not be involved in processes specific to adaptation, the region does appear to be involved in general speech perception processes therefore an effect of TMS would again have been expected.

Whilst no significant effect of TMS was observed in the current study there appears to be a trend towards an effect in Figure 19 and Table 23 with participants in both the STS and PMv group's overall roughly 100 milliseconds slower to respond to the time-compressed speech compared to the no TMS baseline and LOC control condition. In addition, there is a smaller degree of change in response times from the first to the last block in the STS and PMv groups where no reduction in response times is evident, compared to the no TMS and LOC control groups. Indeed, the trend in the STS group is for response times to actually become slower (see Figure 19). Whilst this difference may be due to sampling error, it could also reflect a real difference that is being hidden by the lack of power and therefore sensitivity in the study. Whilst 40 participants is more than many equivalent TMS experiments (see Table 13 in Chapter Three), this equates to just 10 participants per group, and the significant effect of order of condition presentation suggests that each group of 10 is further made up of two subgroups of five participants each. Such a small group size was below the average for equivalent studies and it is therefore possible that the present study lacked enough sensitivity to detect an effect. In order to establish this, more data could be collected for a minimum of an extra five participants per group to make the group sizes in the current study comparable to equivalent experiments.

Despite the lack of a significant effect of TMS, the finding that the order of distortion presentation had a significant effect on both the accuracy and speech of responses for the time-compressed condition is an interesting result and replicates similar findings from Adank and Janse (2009) who found an advantage for perception

of naturally fast speech stimuli if it followed artificially compressed speech and Hervais-Adelman et al. (2011) and Davis et al. (2005) who found a significant benefit in adapting to noise-vocoded words and phrases if a noise-vocoded stimulus followed an undistorted clear speech equivalent. Adank and Janse (2009) frame their result in the context of the RHT (Ahissar et al., 2009) and suggest that the artificially time-compressed signal prepared the perceptual systems for the perception of distorted speech, without stressing the system too much. As a result the perceptual systems were therefore better prepared and primed to adapt to the more extreme naturally compressed distortion, than it would have been had the easier artificial distortion not preceded it. Hervais-Adelman et al. (2011) agree with this conclusion and suggest that the clear speech provides a teaching signal against which the distorted speech can be compared and thus enabling an easier mapping of the distorted sounds onto internal representations. Whilst the natural uncompressed and the distorted compressed sentences were presented in separate blocks in the experiment presented in this chapter, due to the relatively repetitive nature of the stimuli (see Appendix N) it is likely that participants who were tested on the clear sentences first were able to use the knowledge of the repetitive phrase structure to restrict expectations in the time-compressed condition and thus be better able to perform to a higher level in this condition.

Conclusions

In conclusion, this study aimed to determine the extent to which the left ventral premotor cortex is involved in perceptual adaptation to time-compressed speech. Overall, no significant effect of TMS was found at any cortical site of stimulation and as a result possible conclusions regarding the role of the PMv specifically in adaptation to distorted speech stimuli are limited. Whilst it is possible that the wrong cortical sites were targeted, or that the task did not adequately engage the ventral premotor or stress the superior temporal regions for the subtle modulations of TMS to have an effect, it is hypothesised, based on the trend towards significance between the baseline (no TMS and LOC) and experimental conditions (left STS and PMv), that the lack of a significant effect could in fact represent a lack of statistical power in the current study due to the below average number of participants per experimental group. Future

research will have to address the inadequate sample size before more resolute conclusions can be made.

Chapter Five

General Discussion

In this chapter, the experimental work presented in this thesis will be considered in the context of previous research and current neurobiological and cognitive models of speech perception. This chapter will also consider the limitations of the presented research and suggest possible directions for future research.

Summary of research aims

The central aim of this thesis was to characterise the neural and cognitive mechanisms associated with adaptation to distorted speech stimuli with the use of behavioural and Transcranial Magnetic Stimulation methodological paradigms. To this end the following five research aims were addressed:

1. To determine the extent to which learning of one type of speech distortion generalises to the learning of different types of speech distortion (Chapter Two, Experiment One).
2. To expose the underlying cognitive mechanisms associated with individual differences in adaptation profile (Chapter Two, Experiment One).
3. To determine the extent to which exposure to multiple speakers impacts overall adaptation (Chapter Two, Experiments One and Two).
4. To find the most effective TMS protocol to non-invasively impair sentential speech perception in healthy human adults (Chapter Three, Experiments Three to Seven).
5. To investigate the role of the left ventral premotor cortex in adaptation to time-compressed speech (Chapter Four, Experiment Eight).

Review of the originality and impact of presented work and discussion of limitations and future research directions

Generalisation of learning across adverse listening conditions and the underlying cognitive mechanisms (related to the first and second research aims)

The experiments presented in the second chapter of this thesis represent the first set of experiments to investigate adaptation to time-compressed, noise-vocoded and speech in noise in the same set of participants. Their results show that adaptation to one type of speech distortion generalises to other types of distortion. This finding is generally in line with the results in Bent et al. (2016) who investigated the generalisation of learning between a non-native accent, a regional dialect and dysarthric speech and Borrie et al. (2017) who investigated the generalisation from dysarthric speech to/from speech in noise, and is indicative of a general skill within individuals that allows them to adapt to an array of adverse listening conditions. Whilst on first impression this result may seem intuitive, research suggests that adaptation to time-compressed speech occurs at a pre-lexical, phonological level of processing with great importance placed on the isochrony of the language being perceived (Mehler et al., 1993; Pallier et al., 1998; Sebastián-Gallés et al., 2000) whilst adaptation to noise-vocoded speech and speech presented in background noise is believed to be dependent on lexical-semantic access with less importance placed on phonological levels of adaptation (Davis et al., 2005; Hervais-Adelman et al., 2008). This suggests that the brain adopts at least two different approaches when adapting to distorted speech, dependent on the type of distortion, however the results from experiment one suggest that participants possess a general mechanism, which aids adaptation in all conditions irrespective of the distortion. It is possible that the general mechanism involves bidirectional weightings predominately between phonological and lexical-semantic information. For temporal distortions, such as time-compressed speech, phonological adaptations are given more weighting and lexical-semantic information plays a secondary role. Whilst for spectral distortions, such as noise-vocoded speech, adaptation can occur to a limited extent with phonological adaptations (Hervais-Adelman et al., 2008) but lexical-semantic adaptations are more beneficial and therefore are given more weighting.

In addition to the generalisation of learning across conditions in experiment one, the results of measuring individual differences in a battery of audiological and

cognitive mechanisms suggest that overall successful adaptation is dependent on individual differences in vocabulary knowledge, working memory (forward digit span) and general cognitive ability (measured via the MoCA). In Chapter two it was suggested that these three measures represent a form of verbal intelligence that underpins the perceptual learning process and provides support for statistical learning to occur and assists adaptation. This is not the first time that adaptation to distorted speech has been associated with individual differences in vocabulary knowledge or working memory (Banks et al., 2015; Bent et al., 2016; Janse & Adank, 2012; Neger et al., 2014). However, it is the first time that such measures have been linked across distortions (as well as to the undistorted, clear speech condition) in the same set of participants. Both the replication of previous research and the novelty of the within-participant approach adds to the significance of the presented results. Interestingly, performance on the principal component loading most strongly onto vocabulary knowledge, working memory, and general cognition did not predict overall performance in the noise vocoded condition. However, performance on a task involving noise vocoded speech has previously been linked with vocabulary knowledge (Neger et al., 2014). Given the previously discussed association between noise-vocoded speech and lexical-semantic adaptations, it would be expected that individual differences in vocabulary knowledge underpins performance in this form of distortion. Future research might therefore attempt to establish whether a link exists between individual differences in vocabulary knowledge and perceptual learning of noise vocoded speech and if not, which cognitive mechanisms underpin perceptual adaptation to this adverse listening condition.

Limitations and directions for future research

One area of limitation within the first two experiments of this thesis is the lack of a significant change in accuracy of performance in any of the adverse listening conditions. In experiments one and two, adaptation was represented by a significant reduction in response times that were interpreted as participants finding it increasingly easier and therefore becoming quicker to respond accurately with repeated exposure to the adverse listening conditions. Yet the degree of accuracy change was minimal, with less than 10 percent improvement in all conditions, replicating the results of

previous research that have also found minimal changes in accuracy (Adank & Devlin, 2010; Bent et al., 2016; Peelle & Wingfield, 2005). This is particularly relevant as performance failed to reach a ceiling level akin to the clear speech condition in any of the distorted speech conditions, which suggests that further adaptation was possible. As discussed in chapter two, it may be the case that the overall lack of a significant change in accuracy of performance could be due to the distortion parameters used, for example, too few channels used to create the noise vocoded speech resulting in stimuli that was too hard to adapt to. Therefore, future research could address this issue by changing the parameters and observing whether or not greater levels of accuracy could be achieved.

Conversely, it is possible that the lack of a significant change in accuracy is linked to cognitive effort and the willingness of participants to expend excessive amounts of effort on a task for which there is, to them, little incentive to do so (Kool, McGuire, Rosen, & Botvinick, 2010; Westbrook, Kester, & Braver, 2013). Kurzban, Duckworth, Kable, and Myers (2013) argue that most cognitive functions, especially those associated with executive function, have multiple uses for which they can be employed. By using these cognitive functions to perform a certain task, participants experience cognitive effort and fatigue, which Kurzban et al. (2013) argue is related to the participants' desire to attend to and employ the currently used cognitive function for something other than the current task. As a result of the effort required to maintain focus on the present task and not 'redeploy' the cognitive functions elsewhere, task performance plateaus or declines as participants attempt to reduce cognitive strain. It is possible therefore that the lack of a significant level of adaptation (in terms of changes in accuracy) was due to diminishing cognitive effort as trial numbers increased. A limitation therefore of experiments one and two of this thesis is the lack of a measure of cognitive effort, for example, pupillometry (Beatty, 1982; Brown et al., 1999).

Pupillometry refers to the measurement of the size of the pupil and its reactivity in different situations. Changes in pupil size have been strongly associated with cognitive effort, with tasks requiring increased mental activity associated with increased pupillary responses in participants (Beatty, 1982; Kahneman & Beatty, 1966; Piquado, Isaacowitz, & Wingfield, 2010). Whilst initial research investigating the link between pupil dilation and cognitive effort focused on memory load,

subsequent research has shown that pupil size is also affected by language processing at both the individual word (Brown et al., 1999) and sentential level (Hyönä, Tammola, & Alaja, 1995). Most importantly, research has shown that as speech intelligibility decreases (e.g., speech becomes harder to comprehend due to increasing background noise) pupil size increases, suggesting that pupil dilation may represent a valid measure for quantifying listening effort (Kramer, Kapteyn, Festen, & Kuik, 1997; Zekveld, Kramer, & Festen, 2010, 2011). Furthermore, pupillometry can provide a deeper insight into processing load than is possible using just behavioural measures. For example, Koelewijn, Zekveld, Festen, and Kramer (2012) investigated the effect that different background maskers had on SRT performance. Overall, whilst the SNR level differed significantly between maskers indicating that participants could handle greater levels of noise in a single talker masker condition compared to a stationary or fluctuating masker condition before reaching 50 percent correct. The level of cognitive effort required to reach the 50 percent SNR level varied significantly between conditions, with greater cognitive effort required in the single talker condition, as indexed by increased pupil dilation compared to the other masking conditions. This is important as it suggests that mental effort is influenced by more than just the intelligibility of the signal. If mental effort were dependent on the intelligibility then the level of cognitive effort and therefore degree of pupillary response would have been equal across the different masking conditions of Koelewijn et al. (2012) when participants reached the 50 percent correct intelligibility level. This result was replicated by Winn, Edwards, and Litovsky (2015) who found significant individual differences in pupillary dilation amongst participants when only analysing correct response trials. This suggests that to attain the same level of behavioural response requires significant individual differences in cognitive effort. Future research could therefore benefit from the introduction of pupillometry as this data could provide a deeper insight into the level of cognitive effort used/required in the different adverse listening conditions. It is expected that some of the individual differences in overall performance and adaptation rates would be explained by individual differences in cognitive effort.

Additionally, future research could introduce incentives to the participants related to the degree of adaptation that they can achieve in an attempt to reduce the distraction of alternative tasks and maximise cognitive effort throughout the

experimental task. It is possible that with a greater incentive to perform e.g. more course credits or monetary compensation, participants may show overall greater levels of adaptation than observed in the current studies or indeed a different profile of adaptation. It is possible that with all participants more invested in the experimental task and exerting equivalent cognitive effort that the results will better represent individual differences in the actual ability to adapt to the different speech distortions rather than individual differences in engagement with the task.

However, the use of pupillometry as a measure of cognitive effort highlights one of the other limitations of the research of chapter two in that neuropsychological tests and pupillometry represent an indirect measure of cognition which in themselves are still not fully understood. This results in conclusions that are correlational in nature. Yet there are no obvious alternative approaches to measuring the associated underlying cognitive measures responsible for individual differences in adaptation. Therefore, this limitation is inherent to the field of perceptual adaptation and individual differences as a whole, with neuropsychological tests widely used in similar research (for example, see Table 1 in Chapter Two). As a result, the strength of the results is reliant on the replication of findings between experiments as for example the replication of the link between vocabulary knowledge and perception of distorted speech stimuli found in experiment one (Banks et al., 2015; Bent et al., 2016; Janse & Adank, 2012).

Effect of exposure to multiple talkers during adaptation (related to the third research aim)

To investigate whether adaptation to adverse listening conditions was dependent purely on the distortion, the speaker, or an interaction of the two, recordings from multiple speakers were used in all four conditions. In experiment one the talkers were randomly intermixed, whilst in experiment two only one talker was used per distortion. Such an experimental manipulation produced two results of particular interest: (1) despite recording stimuli from four adult males all of roughly the same age and accent and providing no instruction to participants that sentences would be heard from more than one talker, participants appear to have responded to each of the four talkers in a different way. Speaker four on the whole was found to be less intelligible, in both accuracy and response time measures (in both experiments) whilst difficulties in

adapting to speaker three (response times) and speaker two (accuracy) were also found. This supports the notion that adaptation was dependent not only on the condition but also on the speaker. This result in itself is perhaps not overly surprising, outside of the research laboratory it is common to encounter individuals for whom we find it very difficult to comprehend in a noisy environment or an individual whose accent we struggle to adapt to no matter how much listening effort we employ. However, the majority of studies in the field have only used stimuli during the adaptation phase of experiments based on recordings from a single speaker on the assumption that adaptation was specific to the distortion (Adank & Devlin, 2010; Davis et al., 2005) and not dependent on the vocal acoustic characteristics of the speaker. Therefore, despite the intuitive nature of this finding it is important as it shows for the first time that the commonly held assumption that perceptual learning of distorted speech stimuli is dependent on the characteristics of the condition more than the characteristics of the recorded speaker may be inaccurate. Additionally, the finding that the vocal characteristics of the speaker are as important as the acoustic characteristics of the condition has important ramifications beyond the laboratory. Noise-vocoded speech is believed to approximate the experience of using a cochlear implant, and speech occurring in the presence of background noise represents an acoustic environment which individuals find progressively harder to perceive in with increasing age (Gordon-Salant & Fitzgibbons, 1993; Tun, 1998; Tun & Wingfield, 1999; Wong et al., 2009). As hearing loss and changes in the ability to perceive speech are associated with reductions in quality of life (Chia et al., 2007; Dalton et al., 2003), being able to identify which vocal characteristics make a certain individual more or less intelligible could provide an opportunity to develop technology to better assist individuals with hearing impairment. This could potentially result in increased levels of perception and comprehension and overall better quality of life.

The second significant effect of using multiple talkers was that participants performed significantly better in the clear and speech in noise conditions of experiment one where sentences from all four speakers were heard compared to experiment two where the stimuli from just one speaker per condition were heard. This result replicates research in training of foreign/non-native phonemic contrasts that repeatedly show that high variability training leads to a greater level of adaptation and acquisition (Bradlow & Bent, 2008; Shinohara & Iverson, 2018; Wang et al., 1999). It is interesting that

these are the two conditions, in addition to the research on adaptation to accents, that most closely resemble real world situations e.g. when following the conversation of a group of friends in a busy café or listening to a discussion amongst colleagues in a quiet meeting room. It is likely therefore that these are conditions in which participants had vast amounts of experience of perceiving and switching between multiple speakers and therefore held more fine-tuned and appropriate strategies to apply and aid perception. In contrast, attempting to use the same acoustic cues to perceive speech that has been distorted in a non-familiar way, i.e., time-compressed or noise-vocoded resulted in no benefit gained from the high variability training.

Limitations and directions for future research

The results of experiments one and two suggest that the benefits of using highly variable stimuli to assist adaptation to a distorted speech condition may only be experienced in familiar adverse conditions, e.g., clear or accented speech or speech perceived in noise. The benefit does not appear to generalise to all listening conditions. Secondly the results of the first two experiments suggest that adaptation to distorted speech stimuli is dependent on the interaction between the specific characteristics of the distortion and the vocal idiosyncrasies of the speaker. As a result, the most important direction for future research related to the impact of multiple speakers would be to establish which vocal characteristics have the biggest impact on perception and subsequent adaptation. In the experiments of chapter two, participants found speaker four the hardest to perceive and acoustic analyses of each speaker found this speaker to have the lowest median f_0 , whilst speaker one, who was well perceived, produced the overall slowest rate of speech. In future work, the effect of manipulating each of these (and other) vocal characteristics on perceptual adaptation should be investigated. Additionally, it would be interesting to investigate if/how the effect of vocal characteristics of the speaker changes between genders or across the lifespan. For example, children have a significantly higher f_0 level than adults, whilst female adults have a correspondingly higher f_0 than male adults. If the significantly lower f_0 of speaker four was the cause of his reduced intelligibility then it would be hypothesised that the higher f_0 of children and adult females would result in greater levels of intelligibility. Therefore, future research should use stimuli recorded from a more

diverse group of speakers to establish which vocal characteristics impact most severely on perceptual adaptation. Thus far, an extensive amount of research has already been conducted to investigate how different vocal characteristics affect intelligibility. Bradlow, Torretta, and Pisoni (1996) investigated how differences in global measures (i.e., factors that extend over all productions from a specific speaker) and specific pronunciation characteristics (e.g. vowel category realisation) affect intelligibility and found global factors such as gender and age were significant determinants of intelligibility, with female talkers and younger talkers more intelligible than male and older talkers. Conversely on a global level, no effect of speaking rate was found on intelligibility, whilst on a fine-grained, acoustic-phonetic level, Bradlow et al. (1996) found no relationship between average f_0 (independent of gender) and intelligibility but a strong relationship between intelligibility and the range in f_0 and first formant as well as degree of vowel dispersion. Furthermore, Hazan and Baker (2011) investigated the ways in which speakers adjust their acoustic-phonetic characteristics depending on the adverse condition in which they are speaking and found that when the participants' speech was either distorted by a babble masker or was noise-vocoded, participants significantly increased their mean word duration and range of their second formants compared to a clear speech condition. Additionally, whilst, median f_0 , f_0 range and mean energy in the one to three kilohertz range increased significantly in both conditions relative to a clear speech baseline, each of these measures were significantly increased further in the babble masking compared to the vocoded distortion. Future research could aim to investigate what changes occur across more distortions as well as investigating which of these adjustments, if any, are critical to individual differences in intelligibility in adverse listening conditions.

In addition, future research could also focus on speaker identification as opposed to sentence verification and investigate whether participants are able to learn and subsequently identify the four speakers in each of the different speech distortions. If participants are easily able and very accurate at identifying each of the individual speakers across all four conditions then this experimental focus may provide an insight into which idiosyncratic characteristics of the speakers affected perceptual adaptation in experiments one and two of this thesis.

Investigating the most effective TMS protocol to non-invasively impair sentential speech perception in healthy human adults (related to the fourth research aim)

The aim of the research in chapter three was to find the most effective TMS protocol to impair participants' ability to perceive speech in noise. Across the five experiments presented in the chapter, the timing of TMS onset as well as the frequency and intensity of the TMS pulses were adjusted along with the sites of stimulation and control tasks. Overall it was found that the most effective TMS protocol to disrupt processing of aurally presented sentences in noise included the use of a predetermined TMS intensity applied online at a rate of 10Hz starting before the onset and continuing until the end of each sentence. The results of this chapter highlight the potential issues faced and the best practices to adopt when designing TMS experiments with the aim of investigating the neurobiology of speech perception. For example, as shown by Table 13 in Chapter Three, the use of motor thresholds as a measure of cortical responsivity to TMS is becoming the dominant design choice, yet the applicability of motor thresholds to non-motoric cortical regions (or indeed phosphene thresholds to non-visual regions) is yet to be fully established (Stewart et al., 2001; Stokes et al., 2013). Instead the best practice may be to determine the motor threshold of each individual participant and then adjust this intensity depending on the relative depth of the site of stimulation in comparison to the depth of motor cortex, i.e., target sites deeper than motor cortex will require increased intensities whilst sites closer to the cortical surface should result in a reduction in intensity. The efficacy of this proposed procedure could be supported by the results of experiments three and four. These experiments saw no change in behavioural modulation following TMS despite the use of a significantly higher average intensity based on individual motor thresholds in experiment four.

Highlighting the importance of different TMS design choices however is not the most significant finding across the series of experiments presented in chapter three, instead the results of experiment six that used the online rTMS protocol show a TMS-induced impairment in speech perception after stimulation of both left *and* right temporal lobes and thus supports neurobiological models of speech perception that hypothesise bilateral processing in speech perception. These results have important implications for current and future neurobiological models of speech perception.

Limitations and directions for future research

Future research could aim initially to replicate this work before varying the experimental task to establish the role of each hemisphere in speech perception; to establish whether the roles are equivalent and, if not, to establish how they differ (McGettigan & Scott, 2012). For example, the ‘Asymmetric Sampling in Time’ (AST) model from Poeppel (2003) argues that at an early representational level the auditory processing is bilaterally symmetrical. However, beyond this preliminary analysis, the right hemisphere preferentially extracts information over a longer time window (150-250ms; commensurate to syllabic rate of 2-5 kHz) whilst the left hemisphere preferentially extracts information from shorter temporal windows (20-40ms). In contrast, Zatorre and Belin (2001) instead argue that whilst bilateral posterior superior temporal areas respond to temporal variations and bilateral anterior superior temporal regions respond to spectral variations, the weighting of activations is such that the left shows preferences for temporal processing whilst the right is weighted more towards spectral processing. Based on the results of experiments three to seven, hypotheses such as those of Poeppel (2003) and Zatorre and Belin (2001) can be investigated using TMS to establish their validity and investigate whether the different hemispheres perform equivalent or complementary processes. Furthermore, with the use of TMS, future research could also investigate if the roles of the anterior and posterior temporal lobes differ in their contribution to intelligible speech processing. In short, TMS could be used to effectively address some of the remaining questions in the neurobiology of speech perception research.

Role of the left ventral premotor cortex in adaptation to time-compressed speech (related to the fifth research aim)

Experiment eight aimed to determine the extent to which the left ventral premotor cortex is involved in adaptation to time-compressed speech. Beyond its role in the selection and initiation of movements, research thus far has associated the PMv with a task dependent role in speech perception, specifically in adverse listening conditions where the PMv is believed to assist the perceptual systems through additional phonemic segmentation processes. To this end, four groups of participants were asked to perform a computerised auditory version of the SCOLP test for 60 clear,

uncompressed sentences and 60 time-compressed sentences. As the PMv is believed to assist predominantly in adverse listening conditions, an effect on behaviour following cortical modulation through TMS was only expected during perceptual adaptation to the time-compressed speech (i.e., roughly within the first 30 sentences). Once perceptual adaptation had occurred it was believed that this distortion would no longer represent an adverse listening condition and therefore the involvement of the PMv would decline as the more orthodox perceptual regions in temporal lobes become more capable at processing the distorted stimuli without requiring additional input. As a result, no effect of TMS to the PMv was expected in the latter trials of time-compressed speech when processing of this stimuli became more restricted to temporal cortical regions. Overall however, no significant effect of TMS was found at any cortical site of stimulation and as a result possible conclusions of the role of the PMv specifically in adaptation to distorted speech stimuli are limited.

Limitations and directions for future research

Whilst it is possible that the wrong cortical sites were targeted or that the task did not adequately engage the ventral premotor or superior temporal regions for the subtle modulations of TMS to have an effect, it is hypothesised, based on the trend towards a significant difference between the baseline control conditions and the experimental STS and PMv that the lack of a significant effect could in fact be indicative of a lack of statistical power in the current study due to the observed order effects and number of participants per experimental group. Future research could address the inadequate sample size before more resolute conclusions can be made.

Alternatively, functional degeneracy refers to the ability of different cortical structures to perform the same function and/or produce the same outcome. As a result damage to one structure does not result in complete behavioural impairment due to the capabilities of the remaining structures (Price & Friston, 2002). When one structure can fill the functional void left by damage to another structure, the ability to complete a task will be protected. Therefore, it can be hard to ascertain whether or not a damaged cortical region is part of a system that would have been capable of completing the required task had the damage not occurred. It is possible therefore that the system involved in perceptual adaptation to distorted speech stimuli occurs within the confines

of functional degeneracy with multiple structures capable of assisting in the learning process, thus rendering the separate modulation of left PMv or STS insufficient in attempting to disrupt adaptation to time-compressed speech in experiment eight. Future TMS research could therefore potentially investigate this by adopting a ‘condition-and-perturb’ approach to cortical modulation (Andoh et al., 2008; Hartwigsen et al., 2016; O’Shea et al., 2007). This approach combines offline stimulation of one cortical site with online stimulation of another functionally relevant cortical site thereby impairing the functioning in both regions and making it harder for the brain to compensate with the aim of making it easier to observe a behavioural disruption. If perceptual adaptation to distorted speech occurs within a functionally degenerate system it is possible that the condition-and-perturb approach may provide the best opportunity to investigate the different parts of the system.

Conclusions

The majority of everyday communication occurs in the presence of some kind of distortion, yet the human ability to understand speech in adverse listening conditions is remarkably robust. Whilst an extensive amount of research has investigated perceptual adaptation to different speech conditions, our knowledge of the individual differences and the associated cognitive and neural mechanisms affecting perceptual adaptation is still limited. The aim of this thesis therefore was to advance our understanding of this research area. Overall the most significant contributions of this thesis come from the research presented in the second chapter where it was established that individuals possess a general ability to generalise the perceptual learning of one adverse listening condition to others; this perceptual process is underpinned by vocabulary knowledge, working memory and general cognitive ability and is dependent as much on the vocal characteristics of the speaker being perceived as the distortion imposed. In addition, the results from experiment six in chapter three show a TMS-induced impairment in speech (in noise) perception after stimulation of both left and right superior temporal lobes and thus support neurobiological models that hypothesise bilateral processing in speech perception.

Summary of the significant findings of this thesis

1. Perceptual adaptation to one type of speech distortion generalised to other adverse listening conditions
2. Perceptual adaptation was found to be underpinned by a combination of vocabulary knowledge, working memory and general cognitive ability (referred to in this thesis as measures of verbal intelligence).
3. Perceptual adaptation was found to be dependent on both the vocal idiosyncrasies of the speaker as much as the spectral and temporal characteristics of the distortion.
4. Application of online rTMS to bilateral STS significantly impaired perception of sentential speech in noise stimuli.
5. Application of online rTMS to either the left STS or PMv did not significantly impair participants' ability to adapt to time-compressed speech.

Bibliography

- Adank, P. (2012). Design choices in imaging speech comprehension: an Activation Likelihood Estimation (ALE) meta-analysis. *NeuroImage*, 63(3), 1601-1613. doi:10.1016/j.neuroimage.2012.07.027
- Adank, P., Davis, M. H., & Hagoort, P. (2012). Neural dissociation in processing noise and accent in spoken language comprehension. *Neuropsychologia*, 50(1), 77-84. doi:<https://doi.org/10.1016/j.neuropsychologia.2011.10.024>
- Adank, P., & Devlin, J. T. (2010). On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech. *NeuroImage*, 49(1), 1124-1132. doi:<https://doi.org/10.1016/j.neuroimage.2009.07.032>
- Adank, P., & Janse, E. (2009). Perceptual learning of time-compressed and natural fast speech. *J Acoust Soc Am*, 126(5), 2649-2659. doi:10.1121/1.3216914
- Adank, P., & Janse, E. (2010). Comprehension of a novel accent by young and older listeners. *Psychol Aging*, 25(3), 736-740. doi:10.1037/a0020054
- Adank, P., Nuttall, H. E., & Kennedy-Higgins, D. (2017). Transcranial magnetic stimulation and motor evoked potentials in speech perception research. *Language, Cognition and Neuroscience*, 32(7), 900-909. doi:10.1080/23273798.2016.1257816
- Agosta, F., Henry, R. G., Migliaccio, R., Neuhaus, J., Miller, B. L., Dronkers, N. F., . . . Gorno-Tempini, M. L. (2010). Language networks in semantic dementia. *Brain*, 133(1), 286-299. doi:10.1093/brain/awp233
- Ahissar, M., Nahum, M., Nelken, I., & Hochstein, S. (2009). Reverse hierarchies and sensory learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1515), 285-299. doi:10.1098/rstb.2008.0253
- Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *Int J Audiol*, 47 Suppl 2, S53-71. doi:10.1080/14992020802301142
- Alba-Ferrara, L., Ellison, A., & Mitchell, R. L. (2012). Decoding emotional prosody: resolving differences in functional neuroanatomy from fMRI and lesion studies using TMS. *Brain Stimul*, 5(3), 347-353. doi:10.1016/j.brs.2011.06.004
- Amassian, V. E., Cracco, R. Q., Maccabee, P. J., Cracco, J. B., Rudell, A., & Eberle, L. (1989). Suppression of visual perception by magnetic coil stimulation of human occipital cortex. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 74(6), 458-462. doi:[https://doi.org/10.1016/0168-5597\(89\)90036-1](https://doi.org/10.1016/0168-5597(89)90036-1)
- Andoh, J., Artiges, E., Pallier, C., Rivière, D., Mangin, J. F., Cachia, A., . . . Martinot, J. L. (2006). Modulation of language areas with functional MR image-guided magnetic stimulation. *NeuroImage*, 29(2), 619-627. doi:<http://dx.doi.org/10.1016/j.neuroimage.2005.07.029>
- Andoh, J., Artiges, E., Pallier, C., Rivière, D., Mangin, J. F., Paillère-Martinot, M. L., & Martinot, J. L. (2008). Priming Frequencies of Transcranial Magnetic Stimulation over Wernicke's Area Modulate Word Detection. *Cerebral Cortex*, 18(1), 210-216. doi:10.1093/cercor/bhm047
- Andoh, J., & Paus, T. (2011). Combining functional neuroimaging with off-line brain stimulation: modulation of task-related activity in language areas. *J Cogn Neurosci*, 23(2), 349-361. doi:10.1162/jocn.2010.21449

- Ardila, A. (1992). Phonological transformations in conduction aphasia. *Journal of Psycholinguistic Research*, 21(6), 473-484. doi:10.1007/bf01067525
- Baddeley, A., Emslie, H., & Nimmo-Smith, I. (1993). The Spot-the-Word test: a robust estimate of verbal intelligence based on lexical decision. *Br J Clin Psychol*, 32 (Pt 1), 55-65.
- Baddley, A., Emslie, H., & Nimmo-Smith, I. (1992). *The Speech and Capacity of Language Processing Test manual*. Suffolk, UK: Thames Valley Test Company.
- Banai, K., & Lavner, Y. (2012). Perceptual Learning of Time-Compressed Speech: More than Rapid Adaptation. *PLOS ONE*, 7(10). doi:10.1371/journal.pone.0047099
- Banks, B., Gowen, E., Munro, K. J., & Adank, P. (2015). Cognitive predictors of perceptual adaptation to accented speech. *J Acoust Soc Am*, 137(4), 2015-2024. doi:10.1121/1.4916265
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, 91(2), 276-292. doi:10.1037/0033-2909.91.2.276
- Beauchamp, M. S., Nath, A., & Pasalar, S. (2010). fMRI-guided TMS reveals that the STS is a Cortical Locus of the McGurk Effect. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 30(7), 2414-2417. doi:10.1523/JNEUROSCI.4865-09.2010
- Bechtel, W. (2002). Aligning Multiple Research Techniques in Cognitive Neuroscience: Why Is It Important? *Philosophy of Science*, 69(S3), S48-S58. doi:10.1086/341767
- Bent, T., Baese-Berk, M., Borrie, S. A., & McKee, M. (2016). Individual differences in the perception of regional, nonnative, and disordered speech varieties. *The Journal of the Acoustical Society of America*, 140(5), 3775-3786. doi:10.1121/1.4966677
- Bestelmeyer, P. E. G., Belin, P., & Grosbras, M.-H. (2011). Right temporal TMS impairs voice detection. *Current Biology*, 21(20), R838-R839. doi:10.1016/j.cub.2011.08.046
- Bilodeau-Mercure, M., Lortie, C. L., Sato, M., Guitton, M. J., & Tremblay, P. (2015). The neurobiology of speech perception decline in aging. *Brain Struct Funct*, 220(2), 979-997. doi:10.1007/s00429-013-0695-3
- Boatman, D. (2004). Cortical bases of speech perception:evidence from functional lesion studies. *Cognition*, 92(1), 47-65. doi:<https://doi.org/10.1016/j.cognition.2003.09.010>
- Boatman, D., Gordon, B., Hart, J., Selnes, O., Miglioretti, D., & Lenz, F. (2000). Transcortical sensory aphasia: revisited and revised. *Brain*, 123 (Pt 8), 1634-1642.
- Boatman, D., Hart, J., Lesser, R. P., Honeycutt, N., Anderson, N. B., Miglioretti, D., & Gordon, B. (1998). Right hemisphere speech perception revealed by amobarbital injection and electrical interference. *Neurology*, 51(2), 458-464. doi:10.1212/wnl.51.2.458
- Boebinger, D., Evans, S., Rosen, S., Lima, C. F., Manly, T., & Scott, S. K. (2015). Musicians and non-musicians are equally adept at perceiving masked speech. *J Acoust Soc Am*, 137(1), 378-387. doi:10.1121/1.4904537
- Boersma, P., & Weenink, D. (2011). Praat: doing phonetics by computer (Version 5.4.02).

- Borrie, S. A., Baese-Berk, M., Engen, K. V., & Bent, T. (2017). A relationship between processing speech in noise and dysarthric speech. *The Journal of the Acoustical Society of America*, *141*(6), 4660-4667. doi:10.1121/1.4986746
- Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America*, *121*(4), 2339-2349. doi:10.1121/1.2642103
- Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707-729. doi:10.1016/j.cognition.2007.04.005
- Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication*, *20*(3), 255-272. doi:[https://doi.org/10.1016/S0167-6393\(96\)00063-5](https://doi.org/10.1016/S0167-6393(96)00063-5)
- British Society of Audiology, B. (2011). Recommended procedure: Pure-tone air-conduction and bone-conduction threshold audiometry with and without masking. In. <http://www.thebsa.org.uk/resources/>: British Society of Audiology.
- Brown, G. G., Kindermann, S. S., Siegle, G. J., Granholm, E., Wong, E. C., & Buxton, R. B. (1999). Brain activation and pupil response during covert performance of the Stroop Color Word task. *Journal of the International Neuropsychological Society*, *5*(4), 308-319. doi:undefined
- Buchman, A. S., Garron, D. C., Trost-Cardamone, J. E., Wichter, M. D., & Schwartz, M. (1986). Word deafness: one hundred years later. *Journal of Neurology, Neurosurgery & Psychiatry*, *49*(5), 489-499. doi:10.1136/jnnp.49.5.489
- Burk, M. H., Humes, L. E., Amos, N. E., & Strauser, L. E. (2006). Effect of training on word-recognition performance in noise for young normal-hearing and older hearing-impaired listeners. *Ear Hear*, *27*(3), 263-278. doi:10.1097/01.aud.0000215980.21158.a2
- Cainer, K. E., James, C., & Rajan, R. (2008). Learning speech-in-noise discrimination in adult humans. *Hearing Research*, *238*(1), 155-164. doi:<https://doi.org/10.1016/j.heares.2007.10.001>
- Catani, M., Jones, D. K., & ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Ann Neurol*, *57*(1), 8-16. doi:10.1002/ana.20319
- Catani, M., & Thiebaut de Schotten, M. (2012). *Atlas of human brain connections*. New York: Oxford University Press.
- Charpentier, F., & Stella, M. (1986). *Diphone synthesis using an overlap-add technique for speech waveforms concatenation*. Paper presented at the ICASSP '86. IEEE International Conference on Acoustics, Speech, and Signal Processing.
- Chen, R., Classen, J., Gerloff, C., Celnik, P., Wassermann, E. M., Hallett, M., & Cohen, L. G. (1997). Depression of motor cortex excitability by low-frequency transcranial magnetic stimulation. *Neurology*, *48*(5), 1398-1403. doi:10.1212/wnl.48.5.1398
- Chia, E. M., Wang, J. J., Rochtchina, E., Cumming, R. R., Newall, P., & Mitchell, P. (2007). Hearing impairment and health-related quality of life: the Blue Mountains Hearing Study. *Ear Hear*, *28*(2), 187-195. doi:10.1097/AUD.0b013e31803126b6
- Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *J Acoust Soc Am*, *116*(6), 3647-3658.
- Clarke, S., & Morosan, P. (2012). Architecture, Connectivity, and Transmitter Receptors of Human Auditory Cortex. In D. Poeppel, T. Overath, A. N. Popper,

- & R. R. Fay (Eds.), *The Human Auditory Cortex* (pp. 11-38). New York, NY: Springer New York.
- Colenbrander, A. (2002). Visual standards: aspects and ranges of vision loss with emphasis on population surveys. *Report for the International Council of Ophthalmology*, 1-33.
- Counter, S. A., Borg, E., & Lofqvist, L. (1991). Acoustic trauma in extracranial magnetic brain stimulation. *Electroencephalogr Clin Neurophysiol*, 78(3), 173-184.
- Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 14(1), 113-121. doi:10.1037/0096-1523.14.1.113
- Da Costa, S., van der Zwaag, W., Marques, J. P., Frackowiak, R. S. J., Clarke, S., & Saenz, M. (2011). Human Primary Auditory Cortex Follows the Shape of Heschl's Gyrus. *The Journal of Neuroscience*, 31(40), 14067-14075. doi:10.1523/jneurosci.2000-11.2011
- Dalton, D. S., Cruickshanks, K. J., Klein, B. E. K., Klein, R., Wiley, T. L., & Nondahl, D. M. (2003). The Impact of Hearing Loss on Quality of Life in Older Adults. *The Gerontologist*, 43(5), 661-668. doi:10.1093/geront/43.5.661
- Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical Processing in Spoken Language Comprehension. *The Journal of Neuroscience*, 23(8), 3423-3431.
- Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A. G., Taylor, K., & McGettigan, C. (2005). Lexical Information Drives Perceptual Learning of Distorted Speech: Evidence From the Comprehension of Noise-Vocoded Sentences. *Journal of Experimental Psychology: General*, 134(2), 222-241. doi:10.1037/0096-3445.134.2.222
- de Graaf, T. A., Cornelsen, S., Jacobs, C., & Sack, A. T. (2011). TMS effects on subjective and objective measures of vision: stimulation intensity and pre-versus post-stimulus masking. *Conscious Cogn*, 20(4), 1244-1255. doi:10.1016/j.concog.2011.04.012
- de Graaf, T. A., & Sack, A. T. (2011). Null results in TMS: From absence of evidence to evidence of absence. *Neuroscience & Biobehavioral Reviews*, 35(3), 871-877. doi:<https://doi.org/10.1016/j.neubiorev.2010.10.006>
- Deng, Z. D., Lisanby, S. H., & Peterchev, A. V. (2013). Electric field depth-focality tradeoff in transcranial magnetic stimulation: simulation comparison of 50 coil designs. *Brain Stimul*, 6(1), 1-13. doi:10.1016/j.brs.2012.02.005
- Deutsch, G. K., Dougherty, R. F., Bammer, R., Siok, W. T., Gabrieli, J. D., & Wandell, B. (2005). Children's reading performance is correlated with white matter structure measured by diffusion tensor imaging. *Cortex*, 41(3), 354-363.
- DeWitt, I., & Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proc Natl Acad Sci U S A*, 109(8), E505-514. doi:10.1073/pnas.1113427109
- di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Exp Brain Res*, 91(1), 176-180.
- Drager, B., Breitenstein, C., Helmke, U., Kamping, S., & Knecht, S. (2004). Specific and nonspecific effects of transcranial magnetic stimulation on picture-word verification. *Eur J Neurosci*, 20(6), 1681-1687. doi:10.1111/j.1460-9568.2004.03623.x
- Duecker, F., & Sack, A. T. (2015). Rethinking the role of sham TMS. *Frontiers in Psychology*, 6(210). doi:10.3389/fpsyg.2015.00210

- Duffau, H., Gatignol, P., Moritz-Gasser, S., & Mandonnet, E. (2009). Is the left uncinata fasciculus essential for language? *Journal of Neurology*, 256(3), 382. doi:10.1007/s00415-009-0053-9
- Duncan, K. J., Pattamadilok, C., Knierim, I., & Devlin, J. T. (2009). Consistency and variability in functional localisers. *NeuroImage*, 46(4), 1018-1026. doi:<https://doi.org/10.1016/j.neuroimage.2009.03.014>
- Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: effects of talker and rate changes. *J Exp Psychol Hum Percept Perform*, 23(3), 914-927.
- Eckert, M. A., Teubner-Rhodes, S., & Vaden, K. I. J. (2016). Is Listening in Noise Worth It? The Neurobiology of Speech Recognition in Challenging Listening Conditions. *Ear and hearing*, 37, 101S-110S. doi:10.1097/aud.0000000000000300
- Eichele, T., Debener, S., Calhoun, V. D., Specht, K., Engel, A. K., Hugdahl, K., . . . Ullsperger, M. (2008). Prediction of human errors by maladaptive changes in event-related brain networks. *Proc Natl Acad Sci U S A*, 105(16), 6173-6178. doi:10.1073/pnas.0708965105
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech Perception in Infants. *Science*, 171(3968), 303-306. doi:10.1126/science.171.3968.303
- Ellis, R. J., & Munro, K. J. (2013). Does cognitive function predict frequency compressed speech recognition in listeners with normal hearing and normal cognition? *Int J Audiol*, 52(1), 14-22. doi:10.3109/14992027.2012.721013
- Erb, J., Henry, M. J., Eisner, F., & Obleser, J. (2012). Auditory skills and brain morphology predict individual differences in adaptation to degraded speech. *Neuropsychologia*, 50(9), 2154-2164. doi:<http://dx.doi.org/10.1016/j.neuropsychologia.2012.05.013>
- Evans, S. (2012). *The anterior pathway for intelligible speech: insights from univariate and multivariate methods*. UCL (University College London),
- 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. *Cereb Cortex*, 24(9), 2350-2361. doi:10.1093/cercor/bht083
- 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. doi:10.1093/cercor/bht083
- Fairbanks, G., & Jr., F. K. (1957). Word Intelligibility as a Function of Time Compression. *The Journal of the Acoustical Society of America*, 29(5), 636-641. doi:10.1121/1.1908992
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175-191. doi:10.3758/bf03193146
- Fitzgerald, P. B., Fountain, S., & Daskalakis, Z. J. (2006). A comprehensive review of the effects of rTMS on motor cortical excitability and inhibition. *Clin Neurophysiol*, 117(12), 2584-2596. doi:10.1016/j.clinph.2006.06.712
- Formisano, E., Kim, D.-S., Di Salle, F., van de Moortele, P.-F., Ugurbil, K., & Goebel, R. (2003). Mirror-Symmetric Tonotopic Maps in Human Primary Auditory Cortex. *Neuron*, 40(4), 859-869. doi:[https://doi.org/10.1016/S0896-6273\(03\)00669-X](https://doi.org/10.1016/S0896-6273(03)00669-X)

- Friederici, A. D., Kotz, S. A., Scott, S. K., & Obleser, J. (2010). Disentangling syntax and intelligibility in auditory language comprehension. *Human Brain Mapping, 31*(3), 448-457. doi:10.1002/hbm.20878
- Füllgrabe, C., & Rosen, S. (2016). On The (Un)importance of Working Memory in Speech-in-Noise Processing for Listeners with Normal Hearing Thresholds. *Frontiers in Psychology, 7*(1268). doi:10.3389/fpsyg.2016.01268
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119* (Pt 2), 593-609.
- Garvey, W. D. (1953). The intelligibility of speeded speech. *Journal of Experimental Psychology, 45*(2), 102-108. doi:10.1037/h0054381
- Gil-Robles, S., Carvallo, A., Jimenez Mdel, M., Gomez Caicoya, A., Martinez, R., Ruiz-Ocana, C., & Duffau, H. (2013). Double dissociation between visual recognition and picture naming: a study of the visual language connectivity using tractography and brain stimulation. *Neurosurgery, 72*(4), 678-686. doi:10.1227/NEU.0b013e318282a361
- Goldstone, R. L. (1998). Perceptual learning. *Annu Rev Psychol, 49*, 585-612. doi:10.1146/annurev.psych.49.1.585
- Golomb, J. D., Peelle, J. E., & Wingfield, A. (2007). Effects of stimulus variability and adult aging on adaptation to time-compressed speech. *J Acoust Soc Am, 121*(3), 1701-1708.
- Gordon-Salant, S., & Fitzgibbons, P. J. (1993). Temporal Factors and Speech Recognition Performance in Young and Elderly Listeners. *Journal of Speech, Language, and Hearing Research, 36*(6), 1276-1285. doi:10.1044/jshr.3606.1276
- Gordon-Salant, S., Yeni-Komshian, G. H., Fitzgibbons, P. J., Cohen, J. I., & Waldroup, C. (2013). Recognition of accented and unaccented speech in different maskers by younger and older listeners. *The Journal of the Acoustical Society of America, 134*(1), 618-627. doi:10.1121/1.4807817
- Grabski, K., Tremblay, P., Gracco, V. L., Girin, L., & Sato, M. (2013). A mediating role of the auditory dorsal pathway in selective adaptation to speech: a state-dependent transcranial magnetic stimulation study. *Brain research, 1515*, 55-65.
- Green, T., Rosen, S., Faulkner, A., & Paterson, R. (2013). Adaptation to spectrally-rotated speech. *The Journal of the Acoustical Society of America, 134*(2), 1369-1377. doi:10.1121/1.4812759
- Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends in Neurosciences, 25*(7), 348-353. doi:[https://doi.org/10.1016/S0166-2236\(02\)02191-4](https://doi.org/10.1016/S0166-2236(02)02191-4)
- Groppa, S., Oliviero, A., Eisen, A., Quartarone, A., Cohen, L. G., Mall, V., . . . Siebner, H. R. (2012). A practical guide to diagnostic transcranial magnetic stimulation: report of an IFCN committee. *Clin Neurophysiol, 123*(5), 858-882. doi:10.1016/j.clinph.2012.01.010
- Hackett, T. A., Preuss, T. M., & Kaas, J. H. (2001). Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *The Journal of Comparative Neurology, 441*(3), 197-222. doi:10.1002/cne.1407
- Harris, K. C., Dubno, J. R., Keren, N. I., Ahlstrom, J. B., & Eckert, M. A. (2009). Speech Recognition in Younger and Older Adults: A Dependency on Low-Level Auditory Cortex. *The Journal of Neuroscience, 29*(19), 6078-6087. doi:10.1523/jneurosci.0412-09.2009

- Hartwigsen, G., Weigel, A., Schuschan, P., Siebner, H. R., Weise, D., Classen, J., & Saur, D. (2016). Dissociating Parieto-Frontal Networks for Phonological and Semantic Word Decisions: A Condition-and-Perturb TMS Study. *Cerebral Cortex*, *26*(6), 2590-2601. doi:10.1093/cercor/bhv092
- Harvey, D. Y., Wei, T., Ellmore, T. M., Hamilton, A. C., & Schnur, T. T. (2013). Neuropsychological evidence for the functional role of the uncinate fasciculus in semantic control. *Neuropsychologia*, *51*(5), 789-801. doi:<http://dx.doi.org/10.1016/j.neuropsychologia.2013.01.028>
- Hazan, V., & Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *The Journal of the Acoustical Society of America*, *130*(4), 2139-2152. doi:10.1121/1.3623753
- Helfert, R., Snead, C., & Altschuler, R. (1991). The ascending auditory pathways. *Neurobiology of hearing: The central auditory system*, 1-25.
- Hervais-Adelman, A. G., Carlyon, R. P., Johnsrude, I. S., & Davis, M. H. (2012). Brain regions recruited for the effortful comprehension of noise-vocoded words. *Language and Cognitive Processes*, *27*(7-8), 1145-1166. doi:10.1080/01690965.2012.662280
- Hervais-Adelman, A. G., Davis, M. H., Johnsrude, I. S., & Carlyon, R. P. (2008). Perceptual learning of noise vocoded words: Effects of feedback and lexicality. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(2), 460-474. doi:10.1037/0096-1523.34.2.460
- Hervais-Adelman, A. G., Davis, M. H., Johnsrude, I. S., Taylor, K. J., & Carlyon, R. P. (2011). Generalization of perceptual learning of vocoded speech. *J Exp Psychol Hum Percept Perform*, *37*(1), 283-295. doi:10.1037/a0020772
- Hickok, G. (2009). The functional neuroanatomy of language. *Phys Life Rev*, *6*(3), 121-143. doi:10.1016/j.plrev.2009.06.001
- 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. doi:<https://doi.org/10.1016/j.bandl.2008.09.006>
- Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, *4*(4), 131-138. doi:[http://dx.doi.org/10.1016/S1364-6613\(00\)01463-7](http://dx.doi.org/10.1016/S1364-6613(00)01463-7)
- Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*(1), 67-99. doi:<http://dx.doi.org/10.1016/j.cognition.2003.10.011>
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nat Rev Neurosci*, *8*(5), 393-402. doi:10.1038/nrn2113
- Hirnstein, M., Westerhausen, R., & Hugdahl, K. (2013). The Right Planum Temporale Is Involved in Stimulus-Driven, Auditory Attention – Evidence from Transcranial Magnetic Stimulation. *PLOS ONE*, *8*(2), e57316. doi:10.1371/journal.pone.0057316
- Hodges, J. R., & Patterson, K. (2007). Semantic dementia: a unique clinicopathological syndrome. *The Lancet Neurology*, *6*(11), 1004-1014. doi:[https://doi.org/10.1016/S1474-4422\(07\)70266-1](https://doi.org/10.1016/S1474-4422(07)70266-1)
- Hudspeth, A. J. (2014). Integrating the active process of hair cells with cochlear function. *Nature Reviews Neuroscience*, *15*, 600. doi:10.1038/nrn3786

- Humphries, C., Liebenthal, E., & Binder, J. R. (2010). Tonotopic organization of human auditory cortex. *NeuroImage*, *50*(3), 1202-1211. doi:<https://doi.org/10.1016/j.neuroimage.2010.01.046>
- Huyck, J. J., & Johnsrude, I. S. (2012). Rapid perceptual learning of noise-vocoded speech requires attention. *The Journal of the Acoustical Society of America*, *131*(3), EL236-EL242. doi:10.1121/1.3685511
- Hyönä, J., Tommola, J., & Alaja, A. M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *Q J Exp Psychol A*, *48*(3), 598-612.
- Iacoboni, M. (2008). The role of premotor cortex in speech perception: Evidence from fMRI and rTMS. *Journal of Physiology-Paris*, *102*(1), 31-34. doi:<https://doi.org/10.1016/j.jphysparis.2008.03.003>
- IEEE. (1969). IEEE Recommended Practice for Speech Quality Measurements. *IEEE No 297-1969*, 1-24. doi:10.1109/IEEESTD.1969.7405210
- Ilmoniemi, R. J., Ruohonen, J., & Karhu, J. (1999). Transcranial magnetic stimulation--a new tool for functional imaging of the brain. *Crit Rev Biomed Eng*, *27*(3-5), 241-284.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434-446. doi:<https://doi.org/10.1016/j.jml.2007.11.007>
- Janse, E., & Adank, P. (2012). Predicting foreign-accent adaptation in older adults. *The Quarterly Journal of Experimental Psychology*, *65*(8), 1563-1585. doi:10.1080/17470218.2012.658822
- Jeffreys, H. (1961). *The theory of probability* (3rd ed.): OUP Oxford.
- Jung, J., Bungert, A., Bowtell, R., & Jackson, S. R. (2016). Vertex Stimulation as a Control Site for Transcranial Magnetic Stimulation: A Concurrent TMS/fMRI Study. *Brain Stimulation*, *9*(1), 58-64. doi:<http://dx.doi.org/10.1016/j.brs.2015.09.008>
- Kaas, J. H., & Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences*, *97*(22), 11793-11799. doi:10.1073/pnas.97.22.11793
- Kahneman, D., & Beatty, J. (1966). Pupil Diameter and Load on Memory. *Science*, *154*(3756), 1583-1585. doi:10.1126/science.154.3756.1583
- Kaminski, J. A., Korb, F. M., Villringer, A., & Ott, D. V. M. (2011). Transcranial Magnetic Stimulation Intensities in Cognitive Paradigms. *PLOS ONE*, *6*(9), e24836. doi:10.1371/journal.pone.0024836
- Killion, M. C. (2004). Myths about hearing in noise and directional microphones. *Hearing Review*, *11*(2), 14-21.
- Koelwijn, T., Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2012). Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear Hear*, *33*(2), 291-300. doi:10.1097/AUD.0b013e3182310019
- Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: action representation in mirror neurons. *Science*, *297*(5582), 846-848. doi:10.1126/science.1070311
- Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *J Exp Psychol Gen*, *139*(4), 665-682. doi:10.1037/a0020198
- Kramer, S. E., Kapteyn, T. S., Festen, J. M., & Kuik, D. J. (1997). Assessing aspects of auditory handicap by means of pupil dilatation. *Audiology*, *36*(3), 155-164.

- Krieger-Redwood, K., Gaskell, M. G., Lindsay, S., & Jefferies, E. (2013). The selective role of premotor cortex in speech perception: a contribution to phoneme judgements but not speech comprehension. *J Cogn Neurosci*, *25*(12), 2179-2188. doi:10.1162/jocn_a_00463
- Kuhl, P., & Miller, J. (1975). Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. *Science*, *190*(4209), 69-72. doi:10.1126/science.1166301
- Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences*, *36*(6), 661-679. doi:10.1017/S0140525X12003196
- Levitt, H. (1971). Transformed Up-Down Methods in Psychoacoustics. *The Journal of the Acoustical Society of America*, *49*(2B), 467-477. doi:10.1121/1.1912375
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol Rev*, *74*(6), 431-461.
- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1-36. doi:[http://doi.org/10.1016/0010-0277\(85\)90021-6](http://doi.org/10.1016/0010-0277(85)90021-6)
- Loebach, J. L., Pisoni, D. B., & Svirsky, M. A. (2010). Effects of Semantic Context and Feedback on Perceptual Learning of Speech Processed through an Acoustic Simulation of a Cochlear Implant. *Journal of experimental psychology. Human perception and performance*, *36*(1), 224. doi:10.1037/a0017609
- Loizou, P. C., Dorman, M., & Tu, Z. (1999). On the number of channels needed to understand speech. *J Acoust Soc Am*, *106*(4 Pt 1), 2097-2103.
- López-Barroso, D., Catani, M., Ripollés, P., Dell'Acqua, F., Rodríguez-Fornells, A., & de Diego-Balaguer, R. (2013). Word learning is mediated by the left arcuate fasciculus. *Proc Natl Acad Sci U S A*, *110*(32), 13168-13173.
- Maldonado, I. L., Moritz-Gasser, S., & Duffau, H. (2011). Does the left superior longitudinal fascicle subserve language semantics? A brain electrostimulation study. *Brain Struct Funct*, *216*(3), 263-274. doi:10.1007/s00429-011-0309-x
- Mandonnet, E., Nouet, A., Gatignol, P., Capelle, L., & Duffau, H. (2007). Does the left inferior longitudinal fasciculus play a role in language? A brain stimulation study. *Brain*, *130*(Pt 3), 623-629. doi:10.1093/brain/awl361
- Mason, R. A., Prat, C. S., & Just, M. A. (2014). Neurocognitive brain response to transient impairment of Wernicke's area. *Cereb Cortex*, *24*(6), 1474-1484. doi:10.1093/cercor/bhs423
- Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, *27*(7-8), 953-978. doi:10.1080/01690965.2012.705006
- Mayo, L. H., Florentine, M., & Buus, S. (1997). Age of Second-Language Acquisition and Perception of Speech in Noise. *Journal of Speech, Language, and Hearing Research*, *40*(3), 686-693. doi:10.1044/jslhr.4003.686
- McGettigan, C., & Scott, S. K. (2012). Cortical asymmetries in speech perception: what's wrong, what's right and what's left? *Trends in Cognitive Sciences*, *16*(5), 269-276. doi:10.1016/j.tics.2012.04.006
- McGlone, J. (1984). Speech comprehension after unilateral injection of sodium amytal. *Brain and Language*, *22*(1), 150-157. doi:[https://doi.org/10.1016/0093-934X\(84\)90084-1](https://doi.org/10.1016/0093-934X(84)90084-1)

- McShefferty, D., Whitmer, W. M., & Akeroyd, M. A. (2015). The just-noticeable difference in speech-to-noise ratio. *Trends Hear*, *19*. doi:10.1177/2331216515572316
- Mehler, J., Sebastian, N., Altmann, G., Dupoux, E., Christophe, A., & Pallier, C. (1993). Understanding Compressed Sentences: The Role of Rhythm and Meaning. *Ann N Y Acad Sci*, *682*(1), 272-282. doi:10.1111/j.1749-6632.1993.tb22975.x
- Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Curr Biol*, *17*(19), 1692-1696. doi:10.1016/j.cub.2007.08.064
- Mesulam, M. M., Thompson, C. K., Weintraub, S., & Rogalski, E. J. (2015). The Wernicke conundrum and the anatomy of language comprehension in primary progressive aphasia. *Brain*, *138*(8), 2423-2437. doi:10.1093/brain/awv154
- Moerel, M., De Martino, F., & Formisano, E. (2014). An anatomical and functional topography of human auditory cortical areas. *Frontiers in Neuroscience*, *8*, 225. doi:10.3389/fnins.2014.00225
- Moore, B. C., Peters, R. W., & Stone, M. A. (1999). Benefits of linear amplification and multichannel compression for speech comprehension in backgrounds with spectral and temporal dips. *J Acoust Soc Am*, *105*(1), 400-411.
- Moore, D. R., Amitay, S., & Hawkey, D. J. C. (2003). Auditory Perceptual Learning. *Learning & Memory*, *10*(2), 83-85. doi:10.1101/lm.59703
- Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., & Zilles, K. (2001). Human Primary Auditory Cortex: Cytoarchitectonic Subdivisions and Mapping into a Spatial Reference System. *NeuroImage*, *13*(4), 684-701. doi:<https://doi.org/10.1006/nimg.2000.0715>
- Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, *9*(5), 453-467. doi:[https://doi.org/10.1016/0167-6393\(90\)90021-Z](https://doi.org/10.1016/0167-6393(90)90021-Z)
- 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. doi:10.1093/cercor/bhg083
- Nasreddine, Z. S., Phillips, N. A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., . . . Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*, *53*(4), 695-699. doi:10.1111/j.1532-5415.2005.53221.x
- Neger, T. M., Rietveld, T., & Janse, E. (2014). Relationship between perceptual learning in speech and statistical learning in younger and older adults. *Frontiers in Human Neuroscience*, *8*(628). doi:10.3389/fnhum.2014.00628
- Ng, V. K. Y., & Cribbie, R. A. (2017). Using the Gamma Generalized Linear Model for Modeling Continuous, Skewed and Heteroscedastic Outcomes in Psychology. *Current Psychology*, *36*(2), 225-235. doi:10.1007/s12144-015-9404-0
- Niogi, S. N., & McCandliss, B. D. (2006). Left lateralized white matter microstructure accounts for individual differences in reading ability and disability. *Neuropsychologia*, *44*(11), 2178-2188. doi:<http://dx.doi.org/10.1016/j.neuropsychologia.2006.01.011>

- O'Shea, J., Johansen-Berg, H., Trief, D., Göbel, S., & Rushworth, M. F. S. (2007). Functionally Specific Reorganization in Human Premotor Cortex. *Neuron*, 54(3), 479-490. doi:<https://doi.org/10.1016/j.neuron.2007.04.021>
- Obleser, J., Eisner, F., & Kotz, S. A. (2008). Bilateral Speech Comprehension Reflects Differential Sensitivity to Spectral and Temporal Features. *The Journal of Neuroscience*, 28(32), 8116-8123. doi:10.1523/jneurosci.1290-08.2008
- 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. doi:10.1093/cercor/bhp318
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia*, 9, 97-112.
- Osnes, B., Hugdahl, K., & Specht, K. (2011). Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. *NeuroImage*, 54(3), 2437-2445. doi:<https://doi.org/10.1016/j.neuroimage.2010.09.078>
- Pallier, C., Sebastian-Gallés, N., Dupoux, E., Christophe, A., & Mehler, J. (1998). Perceptual adjustment to time-compressed speech: A cross-linguistic study. *Memory & Cognition*, 26(4), 844-851. doi:10.3758/bf03211403
- Palmer, S. D., & Mattys, S. L. (2016). Speech segmentation by statistical learning is supported by domain-general processes within working memory. *The Quarterly Journal of Experimental Psychology*, 69(12), 2390-2401. doi:10.1080/17470218.2015.1112825
- Papagno, C., Miracapillo, C., Casarotti, A., Romero Lauro, L. J., Castellano, A., Falini, A., . . . Bello, L. (2011). What is the role of the uncinate fasciculus? Surgical removal and proper name retrieval. *Brain*, 134(Pt 2), 405-414. doi:10.1093/brain/awq283
- Pascual-Leone, A., Bartres-Faz, D., & Keenan, J. P. (1999). Transcranial magnetic stimulation: studying the brain-behaviour relationship by induction of 'virtual lesions'. *Philos Trans R Soc Lond B Biol Sci*, 354(1387), 1229-1238.
- Peelle, J. E., Johnsrude, I. S., & Davis, M. H. (2010). Hierarchical processing for speech in human auditory cortex and beyond. *Frontiers in Human Neuroscience*, 4(51). doi:10.3389/fnhum.2010.00051
- Peelle, J. E., McMillan, C., Moore, P., Grossman, M., & Wingfield, A. (2004). Dissociable patterns of brain activity during comprehension of rapid and syntactically complex speech: Evidence from fMRI. *Brain and Language*, 91(3), 315-325. doi:<https://doi.org/10.1016/j.bandl.2004.05.007>
- Peelle, J. E., & Wingfield, A. (2005). Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech. *J Exp Psychol Hum Percept Perform*, 31(6), 1315-1330. doi:10.1037/0096-1523.31.6.1315
- Penhune, V. B., Zatorre, R. J., MacDonald, J. D., & Evans, A. C. (1996). Interhemispheric anatomical differences in human primary auditory cortex: probabilistic mapping and volume measurement from magnetic resonance scans. *Cereb Cortex*, 6(5), 661-672.
- Peters, R. W., Moore, B. C., & Baer, T. (1998). Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *J Acoust Soc Am*, 103(1), 577-587.
- Piquado, T., Isaacowitz, D., & Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology*, 47(3), 560-569. doi:doi:10.1111/j.1469-8986.2009.00947.x

- Plack, C. J. (2013). *The sense of hearing*: Psychology Press.
- Plomp, R., & Mimpen, A. M. (1979a). Improving the Reliability of Testing the Speech Reception Threshold for Sentences. *Audiology*, 18(1), 43-52. doi:10.3109/00206097909072618
- Plomp, R., & Mimpen, A. M. (1979b). Speech-reception threshold for sentences as a function of age and noise level. *J Acoust Soc Am*, 66(5), 1333-1342.
- Plomp, R., & Mimpen, A. M. (1979c). Speech-reception threshold for sentences as a function of age and noise level. *The Journal of the Acoustical Society of America*, 66(5), 1333-1342. doi:10.1121/1.383554
- Poepfel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication*, 41(1), 245-255. doi:[https://doi.org/10.1016/S0167-6393\(02\)00107-3](https://doi.org/10.1016/S0167-6393(02)00107-3)
- Poldrack, R. A., Temple, E., Protopapas, A., Nagarajan, S., Tallal, P., Merzenich, M., & Gabrieli, J. D. E. (2001). Relations between the Neural Bases of Dynamic Auditory Processing and Phonological Processing: Evidence from fMRI. *J Cogn Neurosci*, 13(5), 687-697. doi:10.1162/089892901750363235
- Price, C. J., & Friston, K. J. (2002). Degeneracy and cognitive anatomy. *Trends in Cognitive Sciences*, 6(10), 416-421. doi:[https://doi.org/10.1016/S1364-6613\(02\)01976-9](https://doi.org/10.1016/S1364-6613(02)01976-9)
- Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci*, 12(6), 718-724. doi:http://www.nature.com/neuro/journal/v12/n6/supinfo/nn.2331_S1.html
- Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Brain Res Cogn Brain Res*, 3(2), 131-141.
- Robles, L., & Ruggero, M. A. (2001). Mechanics of the Mammalian Cochlea. *Physiological Reviews*, 81(3), 1305-1352. doi:10.1152/physrev.2001.81.3.1305
- Romero, J. R., Ansel, D., Sparing, R., Gangitano, M., & Pascual-Leone, A. (2002). Subthreshold low frequency repetitive transcranial magnetic stimulation selectively decreases facilitation in the motor cortex. *Clinical Neurophysiology*, 113(1), 101-107. doi:[http://dx.doi.org/10.1016/S1388-2457\(01\)00693-9](http://dx.doi.org/10.1016/S1388-2457(01)00693-9)
- Rönnerberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., . . . Rudner, M. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience*, 7(31). doi:10.3389/fnsys.2013.00031
- Rönnerberg, J., Rudner, M., Foo, C., & Lunner, T. (2008). Cognition counts: a working memory system for ease of language understanding (ELU). *Int J Audiol*, 47 Suppl 2, S99-105. doi:10.1080/14992020802301167
- Rosen, S., & Howell, P. (2011). *Signals and Systems for Speech and Hearing*: Brill Academic Pub.
- Rosen, S., Wise, R. J. S., Chadha, S., Conway, E., & Scott, S. K. (2011). Hemispheric Asymmetries in Speech Perception: Sense, Nonsense and Modulations. *PLOS ONE*, 6(9), e24672. doi:10.1371/journal.pone.0024672
- Rossi, S., Hallett, M., Rossini, P. M., & Pascual-Leone, A. (2009). Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clinical Neurophysiology*, 120(12), 2008-2039. doi:<https://doi.org/10.1016/j.clinph.2009.08.016>

- Rossini, P. M., Barker, A. T., Berardelli, A., Caramia, M. D., Caruso, G., Cracco, R. Q., . . . Tomberg, C. (1994). Non-invasive electrical and magnetic stimulation of the brain, spinal cord and roots: basic principles and procedures for routine clinical application. Report of an IFCN committee. *Electroencephalography and Clinical Neurophysiology*, *91*(2), 79-92. doi:10.1016/0013-4694(94)90029-9
- Rossini, P. M., Burke, D., Chen, R., Cohen, L. G., Daskalakis, Z., Di Iorio, R., . . . Ziemann, U. (2015). Non-invasive electrical and magnetic stimulation of the brain, spinal cord, roots and peripheral nerves: Basic principles and procedures for routine clinical and research application. An updated report from an I.F.C.N. Committee. *Clin Neurophysiol*, *126*(6), 1071-1107. doi:10.1016/j.clinph.2015.02.001
- Sadaghiani, S., & D'Esposito, M. (2015). Functional Characterization of the Cingulo-Opercular Network in the Maintenance of Tonic Alertness. *Cerebral Cortex*, *25*(9), 2763-2773. doi:10.1093/cercor/bhu072
- Sato, M., Tremblay, P., & Gracco, V. L. (2009). A mediating role of the premotor cortex in phoneme segmentation. *Brain and Language*, *111*(1), 1-7. doi:<https://doi.org/10.1016/j.bandl.2009.03.002>
- 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*, *105*(46), 18035-18040. doi:10.1073/pnas.0805234105
- Schnupp, J., Nelken, I., & King, A. (2011). *Auditory neuroscience: Making sense of sound*: MIT press.
- Schubotz, R. I., & von Cramon, D. Y. (2003). Functional-anatomical concepts of human premotor cortex: evidence from fMRI and PET studies. *NeuroImage*, *20 Suppl 1*, S120-131.
- Schwartz, J.-L., Basirat, A., Ménard, L., & Sato, M. (2012). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, *25*(5), 336-354. doi:<https://doi.org/10.1016/j.jneuroling.2009.12.004>
- Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123*(12), 2400-2406. doi:10.1093/brain/123.12.2400
- Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, *26*(2), 100-107. doi:[https://doi.org/10.1016/S0166-2236\(02\)00037-1](https://doi.org/10.1016/S0166-2236(02)00037-1)
- Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action — candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience*, *10*, 295. doi:10.1038/nrn2603
- Scott, S. K., Rosen, S., Lang, H., & Wise, R. J. S. (2006). Neural correlates of intelligibility in speech investigated with noise vocoded speech—A positron emission tomography study. *The Journal of the Acoustical Society of America*, *120*(2), 1075-1083. doi:10.1121/1.2216725
- Sebastián-Gallés, N., Dupoux, E., Costa, A., & Mehler, J. (2000). Adaptation to time-compressed speech: Phonological determinants. *Perception & Psychophysics*, *62*(4), 834-842. doi:10.3758/bf03206926
- Selnes, O. A., Knopman, D. S., Niccum, N., & Rubens, A. B. (1985). The critical role Wernicke's area in sentence repetition. *Annals of Neurology*, *17*(6), 549-557. doi:10.1002/ana.410170604

- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270(5234), 303-304.
- Shinohara, Y., & Iverson, P. (2018). High variability identification and discrimination training for Japanese speakers learning English /r/-/l/. *Journal of Phonetics*, 66, 242-251. doi:<https://doi.org/10.1016/j.wocn.2017.11.002>
- Shtyrov, Y., Kujala, T., Ahveninen, J., Tervaniemi, M., Alku, P., Ilmoniemi, R. J., & Naatanen, R. (1998). Background acoustic noise and the hemispheric lateralization of speech processing in the human brain: magnetic mismatch negativity study. *Neurosci Lett*, 251(2), 141-144.
- Silvanto, J., & Muggleton, N. G. (2008). New light through old windows: Moving beyond the “virtual lesion” approach to transcranial magnetic stimulation. *NeuroImage*, 39(2), 549-552. doi:<https://doi.org/10.1016/j.neuroimage.2007.09.008>
- Slevc, R. L., & Shell, A. R. (2015). Chapter 32 - Auditory agnosia. In M. J. Aminoff, F. Boller, & D. F. Swaab (Eds.), *Handbook of Clinical Neurology* (Vol. 129, pp. 573-587): Elsevier.
- Sliwiska, M. W., James, A., & Devlin, J. T. (2015). Inferior Parietal Lobule Contributions to Visual Word Recognition. *J Cogn Neurosci*, 27(3), 593-604. doi:10.1162/jocn_a_00721
- Sliwiska, M. W., Khadilkar, M., Campbell-Ratcliffe, J., Quevenco, F., & Devlin, J. T. (2012). Early and Sustained Supramarginal Gyrus Contributions to Phonological Processing. *Frontiers in Psychology*, 3(161). doi:10.3389/fpsyg.2012.00161
- Sliwiska, M. W., Vitello, S., & Devlin, J. T. (2014). Transcranial Magnetic Stimulation for Investigating Causal Brain-behavioral Relationships and their Time Course. *Journal of Visualized Experiments : JoVE*(89), 51735. doi:10.3791/51735
- Song, J. H., Skoe, E., Banai, K., & Kraus, N. (2012). Training to Improve Hearing Speech in Noise: Biological Mechanisms. *Cerebral Cortex*, 22(5), 1180-1190. doi:10.1093/cercor/bhr196
- Spencer, M., Kaschak, M. P., Jones, J. L., & Lonigan, C. J. (2015). Statistical learning is related to early literacy-related skills. *Reading and Writing*, 28(4), 467-490. doi:10.1007/s11145-014-9533-0
- Stassenko, A., Garcea, F. E., & Mahon, B. Z. (2013). What happens to the motor theory of perception when the motor system is damaged? *Language and cognition*, 5(2-3), 225-238.
- Stewart, L. M., Walsh, V., & Rothwell, J. C. (2001). Motor and phosphene thresholds: a transcranial magnetic stimulation correlation study. *Neuropsychologia*, 39(4), 415-419.
- Stokes, M. G., Barker, A. T., Dervinis, M., Verbruggen, F., Maizey, L., Adams, R. C., & Chambers, C. D. (2013). Biophysical determinants of transcranial magnetic stimulation: effects of excitability and depth of targeted area. *J Neurophysiol*, 109(2), 437-444. doi:10.1152/jn.00510.2012
- Tao, L., & Taft, M. (2017). Influences of Cognitive Processing Capacities on Speech Perception in Young Adults. *Frontiers in Psychology*, 8(266). doi:10.3389/fpsyg.2017.00266
- Thiebaut de Schotten, M., ffytche, D. H., Bizzi, A., Dell'Acqua, F., Allin, M., Walshe, M., . . . Catani, M. (2011). Atlasing location, asymmetry and inter-subject variability of white matter tracts in the human brain with MR diffusion

- tractography. *NeuroImage*, 54(1), 49-59.
doi:<http://dx.doi.org/10.1016/j.neuroimage.2010.07.055>
- Tomasino, B., Marin, D., Maieron, M., D'Agostini, S., Medeossi, I., Fabbro, F., . . . Luzzatti, C. (2015). A multimodal mapping study of conduction aphasia with impaired repetition and spared reading aloud. *Neuropsychologia*, 70, 214-226. doi:<https://doi.org/10.1016/j.neuropsychologia.2015.02.023>
- Tremblay, P., & Small, S. L. (2011). On the context-dependent nature of the contribution of the ventral premotor cortex to speech perception. *NeuroImage*, 57(4), 1561-1571. doi:10.1016/j.neuroimage.2011.05.067
- Tun, P. A. (1998). Fast noisy speech: Age differences in processing rapid speech with background noise. *Psychol Aging*, 13(3), 424-434. doi:10.1037/0882-7974.13.3.424
- Tun, P. A., & Wingfield, A. (1999). One voice too many: adult age differences in language processing with different types of distracting sounds. *J Gerontol B Psychol Sci Soc Sci*, 54(5), P317-327.
- Vaden, K. I., Kuchinsky, S. E., Ahlstrom, J. B., Dubno, J. R., & Eckert, M. A. (2015). Cortical Activity Predicts Which Older Adults Recognize Speech in Noise and When. *The Journal of Neuroscience*, 35(9), 3929-3937. doi:10.1523/jneurosci.2908-14.2015
- Vaden, K. I., Kuchinsky, S. E., Cute, S. L., Ahlstrom, J. B., Dubno, J. R., & Eckert, M. A. (2013). The Cingulo-Opercular Network Provides Word-Recognition Benefit. *The Journal of Neuroscience*, 33(48), 18979-18986. doi:10.1523/jneurosci.1417-13.2013
- Versfeld, N. J., & Dreschler, W. A. (2002). The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners. *J Acoust Soc Am*, 111(1 Pt 1), 401-408.
- Vongphoe, M., & Zeng, F.-G. (2005). Speaker recognition with temporal cues in acoustic and electric hearing. *The Journal of the Acoustical Society of America*, 118(2), 1055-1061. doi:10.1121/1.1944507
- Voor, J. B., & Miller, J. M. (1965). The effect of practice upon the comprehension of time-compressed speech. *Communications Monographs*, 32(4), 452-454.
- Wada, J. (1949). A new method of determining the side of cerebral speech dominance: a preliminary report on the intracarotid injection of sodium amytal in man. *Igaku to Seibutsugaku*, 14, 221-222.
- Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, 106(6), 3649-3658. doi:10.1121/1.428217
- 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. doi:<https://doi.org/10.1016/j.tins.2005.09.010>
- Watson, C. S. (1980). Time Course of Auditory Perceptual Learning. *Annals of Otolaryngology, Rhinology & Laryngology*, 89(5_suppl), 96-102. doi:10.1177/00034894800890s525
- Weissman, D. H., Roberts, K. C., Visscher, K. M., & Woldorff, M. G. (2006). The neural bases of momentary lapses in attention. *Nat Neurosci*, 9, 971. doi:10.1038/nn1727
- Wessinger, C. M., Buonocore, M. H., Kussmaul, C. L., & Mangun, G. R. (1997). Tonotopy in human auditory cortex examined with functional magnetic resonance imaging. *Human Brain Mapping*, 5(1), 18-25. doi:10.1002/(SICI)1097-0193(1997)5:1<18::AID-HBM3>3.0.CO;2-Q

- Wessinger, C. M., VanMeter, J., Tian, B., Van Lare, J., Pekar, J., & Rauschecker, J. P. (2001). Hierarchical Organization of the Human Auditory Cortex Revealed by Functional Magnetic Resonance Imaging. *J Cogn Neurosci*, *13*(1), 1-7. doi:10.1162/089892901564108
- Westbrook, A., Kester, D., & Braver, T. S. (2013). What Is the Subjective Cost of Cognitive Effort? Load, Trait, and Aging Effects Revealed by Economic Preference. *PLOS ONE*, *8*(7). doi:<http://dx.doi.org/10.1371/journal.pone.0068210>
- Whitmer, W. M., McShefferty, D., & Akeroyd, M. A. (2016). *On Detectable and Meaningful Speech-Intelligibility Benefits*, Cham.
- Wilson, S. M., & Iacoboni, M. (2006). Neural responses to non-native phonemes varying in producibility: Evidence for the sensorimotor nature of speech perception. *NeuroImage*, *33*(1), 316-325. doi:<https://doi.org/10.1016/j.neuroimage.2006.05.032>
- Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nat Neurosci*, *7*, 701. doi:10.1038/nn1263
- <https://www.nature.com/articles/nn1263#supplementary-information>
- Winn, M. B., Edwards, J. R., & Litovsky, R. Y. (2015). The Impact of Auditory Spectral Resolution on Listening Effort Revealed by Pupil Dilation. *Ear and hearing*, *36*(4), e153-e165. doi:10.1097/AUD.000000000000145
- Wise, S. P. (1985). The Primate Premotor Cortex: Past, Present, and Preparatory. *Annual Review of Neuroscience*, *8*(1), 1-19. doi:10.1146/annurev.ne.08.030185.000245
- Wong, P. C., Jin, J. X., Gunasekera, G. M., Abel, R., Lee, E. R., & Dhar, S. (2009). Aging and cortical mechanisms of speech perception in noise. *Neuropsychologia*, *47*(3), 693-703. doi:10.1016/j.neuropsychologia.2008.11.032
- Wong, P. C., Uppunda, A. K., Parrish, T. B., & Dhar, S. (2008). Cortical mechanisms of speech perception in noise. *J Speech Lang Hear Res*, *51*(4), 1026-1041. doi:10.1044/1092-4388(2008/075)
- Yeatman, J. D., Dougherty, R. F., Rykhlevskaia, E., Sherbondy, A. J., Deutsch, G. K., Wandell, B. A., & Ben-Shachar, M. (2011). Anatomical Properties of the Arcuate Fasciculus Predict Phonological and Reading Skills in Children. *J Cogn Neurosci*, *23*(11), 3304-3317.
- Young, E. D. (2008). Neural representation of spectral and temporal information in speech. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 923-945. doi:10.1098/rstb.2007.2151
- Yousry, T. A., Schmid, U. D., Alkadhi, H., Schmidt, D., Peraud, A., Buettner, A., & Winkler, P. (1997). Localization of the motor hand area to a knob on the precentral gyrus. A new landmark. *Brain*, *120* (Pt 1), 141-157.
- Zaballos, M. T., Plasencia, D. P., Gonzalez, M. L., de Miguel, A. R., & Macias, A. R. (2016). Air traffic controllers' long-term speech-in-noise training effects: A control group study. *Noise Health*, *18*(85), 376-381. doi:10.4103/1463-1741.195804
- Zatorre, R. J., & Belin, P. (2001). Spectral and Temporal Processing in Human Auditory Cortex. *Cerebral Cortex*, *11*(10), 946-953. doi:10.1093/cercor/11.10.946

- Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., & Schoonhoven, R. (2006). Top-down and bottom-up processes in speech comprehension. *NeuroImage*, 32(4), 1826-1836. doi:<https://doi.org/10.1016/j.neuroimage.2006.04.199>
- Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful listening: the influence of sentence intelligibility. *Ear Hear*, 31(4), 480-490. doi:10.1097/AUD.0b013e3181d4f251
- Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011). Cognitive load during speech perception in noise: the influence of age, hearing loss, and cognition on the pupil response. *Ear Hear*, 32(4), 498-510. doi:10.1097/AUD.0b013e31820512bb

List of Appendices

Appendix A	211
List of IEEE sentences used for the assessment of Speech Reception Threshold in Experiment One and Eight.	211
Appendix B	213
List for forward digit span test of working memory used in Experiment One. ...	213
Appendix C	214
SCOLP Spot the Word Vocabulary Knowledge Test.	214
Appendix D	215
Definition of the correct answers to the SCOLP Spot the Word Vocabulary Knowledge Test.....	215
Appendix E	216
Parts A and B of the Trail Making Test as used in Experiment One.	216
Part A.....	216
Part B.....	216
Appendix F	217
Version A of the Montreal Cognitive Assessment (MoCA) used in Experiment One.	217
Appendix G	218
Lists of Speed and Capacity of Language Processing (SCOLP) sentences used in Experiments One and Two.....	218
Appendix H	219
Full list of contrast follow-up results for the generalised linear mixed model, accuracy and response times data for Experiment One.....	219
Accuracy.....	219
Response Times.....	220
Appendix I	223

Full list of contrast follow-up results for the generalised linear mixed model, accuracy and response times data for Experiment Two.....	223
Accuracy	223
Response Times	224
Appendix J.....	226
List of IEEE sentences used for the assessment of Speech Reception Threshold in Experiment Three, Four and Five.	226
List A.....	226
List B.....	226
List C.....	227
List D.....	228
Appendix K.....	230
List of IEEE sentences used for the assessment of Speech Reception Threshold in Experiment Six.....	230
List A.....	230
List B.....	230
List C.....	231
List D.....	232
Appendix L	234
Lists of Speed and Capacity of Language Processing (SCOLP) sentences used in Experiment Seven to determine reading time.	234
Appendix M.....	235
List of scrambled non-word sentences used the visual discrimination task of Experiment Seven.	235
List A.....	235
List B.....	235
List C.....	236
List D.....	237

Appendix N 239

Lists of Speed and Capacity of Language Processing (SCOLP) sentences used in
Experiment Eight..... 239

Appendix B

List for forward digit span test of working memory used in Experiment One.

3	829	683	841
	132	356	251
	Total out of 6 =		
4	6241	1362	5316
	2359	8392	4815
	Total out of 6 =		
5	84132	85293	19514
	62143	91635	82691
	Total out of 6 =		
6	584261	492615	148239
	261384	246681	423896
	Total out of 6 =		
7	2941318	6292865	1893562
	1285394	8243164	3185624
	Total out of 6 =		
8	65148299	28653190	85129136
	18442913	65392381	26591243
	Total out of 6 =		
9	619124382	239894615	539048216
	346231958	864934612	513985265
	Total out of 6 =		
10	4982106453	2853961624	2914984352
	5631298426	9381434826	6983285149
	Total out of 6 =		
	Digit Span =		

Appendix C

SCOLP Spot the Word Vocabulary Knowledge Test.

Real words highlighted in bold font.

REMOVED TO AVOID COPYRIGHT INFRINGEMENT

See the following sources for relevant information:

Baddeley, A., Emslie, H., & Nimmo-Smith, I. (1993). The Spot-the-Word test: a robust estimate of verbal intelligence based on lexical decision. *Br J Clin Psychol*, 32 (1), 55-65.

Baddeley, A., Emslie, H., & Nimmo-Smith, I. (1992). *The Speech and Capacity of Language Processing Test manual*. Suffolk, UK: Thames Valley Test Company.

Appendix D

Definition of the correct answers to the SCOLP Spot the Word Vocabulary Knowledge Test.

REMOVED TO AVOID COPYRIGHT INFRINGEMENT

See the following sources for relevant information:

Baddeley, A., Emslie, H., & Nimmo-Smith, I. (1993). The Spot-the-Word test: a robust estimate of verbal intelligence based on lexical decision. *Br J Clin Psychol*, 32 (1), 55-65.

Baddley, A., Emslie, H., & Nimmo-Smith, I. (1992). *The Speech and Capacity of Language Processing Test manual*. Suffolk, UK: Thames Valley Test Company.

Appendix E

Parts A and B of the Trail Making Test as used in Experiment One.

Part A

REMOVED TO AVOID COPYRIGHT INFRINGEMENT

Part B

REMOVED TO AVOID COPYRIGHT INFRINGEMENT

Appendix F

Version A of the Montreal Cognitive Assessment (MoCA) used in Experiment One.

REMOVED TO AVOID COPYRIGHT INFRINGEMENT

See the following sources for relevant information:

Nasreddine, Z. S., Phillips, N. A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., . . . Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*, *53*(4), 695-699.
doi:10.1111/j.1532-5415.2005.53221.x

Appendix G

Lists of Speed and Capacity of Language Processing (SCOLP) sentences used in Experiments One and Two

REMOVED TO AVOID COPYRIGHT INFRINGEMENT

See the following sources for relevant information:

Baddeley, A., Emslie, H., & Nimmo-Smith, I. (1993). The Spot-the-Word test: a robust estimate of verbal intelligence based on lexical decision. *Br J Clin Psychol*, 32 (1), 55-65.

Baddley, A., Emslie, H., & Nimmo-Smith, I. (1992). *The Speech and Capacity of Language Processing Test manual*. Suffolk, UK: Thames Valley Test Company.

Appendix H

Full list of contrast follow-up results for the generalised linear mixed model, accuracy and response times data for Experiment One

Accuracy

Formula: Accuracy ~ 1 + Block_trial_no + Condition + Speaker + (1|Participant)

ANOVA:

Condition: $F(3,17272) = 136.31, p < 0.001$

Speaker: $F(3,17272) = 6.4, p = 0.0002$

Contrasts:

Conditions:

Speech in Noise vs Vocoded: $F(1,17272) = 29.82, p < 0.001$

Speech in Noise vs Time Compressed: $F(1,17272) = 65.72, p < 0.001$

Vocoded vs Time Compressed: $F(1,17272) = 181.87, p < 0.001$

Condition	Clear	Time Compressed	Vocoded	Speech in Noise
Mean	95.79	85.37	60.65	70.09
Std Dev.	20.09	35.34	48.85	45.79
Conf. Int.	[95.19, 96.39]	[84.32, 86.42]	[59.19, 62.11]	[68.73, 71.46]

Speaker:

Two vs Three: $F(1,17272) = 0.49, p = 0.48$

Two vs Four: $F(1,17272) = 16.93, p < 0.001$

Three vs Four: $F(1,17272) = 11.74, p < 0.001$

Speaker	One	Two	Three	Four
Mean	78.26	81.16	79.42	73.06
Std Dev.	41.25	39.11	40.43	44.37
Conf. Int.	[77.03, 79.49]	[79.99, 82.32]	[78.22, 80.63]	[71.73, 74.38]

Response Times

Formula: $RT \sim 1 + \text{Block_trial_no} * \text{Condition} + \text{Block_trial_no} * \text{Speaker} + \text{Condition} * \text{Speaker} + \text{Block_trial_no} * \text{Condition} * \text{Speaker} + (1 | \text{Participant})$

ANOVA:

Condition: $F(3,13442) = 35.59, p < 0.001$

Speaker: $F(3,13442) = 4.67, p = 0.002$

Block_trial_no*Speaker: $F(3,13442) = 5.8, p = 0.0005$

Block_trial_no*Condition*Speaker: $F(9, 13442) = 2.96, p = 0.001$

Contrasts:

Conditions:

Speech in Noise vs Vocoded: $F(1,13442) = 12.56, p < 0.001$

Speech in Noise vs Time Compressed: $F(1, 13442) = 19.42, p < 0.001$

Vocoded vs Time Compressed: $F(1, 13442) = 0.31, p = 0.57$

Condition	Clear	Time Compressed	Vocoded	Speech in Noise
Mean	409.97	818.99	743.29	611.46
Std Dev.	361.39	435.55	458.43	455.32
Conf. Int.	[398.96, 420.99]	[804.93, 833.05]	[725.73, 760.86]	[595.24, 627.69]

Speaker:

Two vs Three: $F(1, 13442) = 7.09, p = 0.007$

Two vs Four: $F(1, 13442) = 9.59, p = 0.002$

Three vs Four: $F(1, 13442) = 0.91, p = 0.33$

Speaker	One	Two	Three	Four
Mean	544.69	616.35	682.21	688.43
Std Dev.	446.3	437.76	457.71	432.71
Conf. Int.	[529.64, 559.74]	[601.85, 630.84]	[666.89, 697.53]	[672.28, 704.58]

Block_trial_no *Speaker:

Trial_no/Sp Two vs Trial_no/Sp Three: $F(1,13442) = 5.99, p = 0.01$

Trial_no/Sp Two vs Trial_no/Sp Four: $F(1,13442) = 8.59, p = 0.003$

Trial_no/Sp Three vs Trial_no/Sp Four: $F(1,13442) = 0.7, p = 0.4$

Block_trial_no*Condition*Speaker:

Trial_no/Noise/Sp Two vs Trial_no/Noise/Sp Three: $F(1,13442) = 6.7, p = 0.009$

Trial_no/Noise/Sp Two vs Trial_no/Noise/Sp Four: $F(1,13442) = 4.99, p = 0.02$

Trial_no/Noise/Sp Three vs Trial_no/Noise/Sp Four: $F(1,13442) < 0.1, p = 0.99$

Trial_no/Vocoded/Sp Two vs Trial_no/Vocoded/Sp Three: $F(1,13442) = 6.56, p = 0.01$

Trial_no/Vocoded/Sp Two vs Trial_no/Vocoded/Sp Four: $F(1,13442) = 2.26, p = 0.13$

Trial_no/Vocoded/Sp Three vs Trial_no/Vocoded/Sp Four: $F(1,13442) = 0.52, p = 0.46$

Trial_no/Compressed/S Two vs Trial_no/ Compressed /S Three: $F(1,13442) = 4.19, p = 0.04$

Trial_no/Compressed/S Two vs Trial_no/ Compressed /S Three: $F(1,13442) = 4.01, p = 0.04$

Trial_no/Compressed/S Two vs Trial_no/ Compressed /S Three: $F(1,13442) = 0.02$,
 $p = 0.87$

Appendix I

Full list of contrast follow-up results for the generalised linear mixed model, accuracy and response times data for Experiment Two

Accuracy

Formula: Accuracy ~ 1 + Block_trial_no + Condition + Speaker + (1|Participant)

ANOVA:

Condition: $F(3, 4600) = 47.53, p < 0.001$

Contrasts:

Conditions:

Speech in Noise vs Vocoded: $F(1, 4600) = 6.97, p = 0.009$

Speech in Noise vs Time Compressed: $F(1, 4600) = 34.91, p < 0.001$

Vocoded vs Time Compressed: $F(1, 4600) = 71.08, p < 0.001$

Condition	Clear	Time Compressed	Vocoded	Speech in Noise
Mean	92.47	81.82	53.58	63.42
Std Dev.	26.39	38.58	49.89	48.19
Conf. Int.	[90.71, 94.23]	[79.25, 84.4]	[50.25, 56.91]	[60.2, 66.64]

Response Times

Formula: $RT \sim 1 + \text{Block_trial_no} * \text{Condition} + \text{Block_trial_no} * \text{Speaker} + \text{Condition} * \text{Speaker} + (1 | \text{Participant})$

ANOVA:

Condition: $F(3, 3356) = 6.88, p = 0.0001$

Speaker: $F(3, 3356) = 2.97, p = 0.03$

Block_trial_no*Condition: $F(3, 3356) = 4.1, p = 0.006$

Condition*Speaker: $F(3, 3356) = 5.08, p < 0.001$

Contrasts:

Conditions:

Speech in Noise vs Vocoded: $F(1, 3356) = 1.33, p = 0.24$

Speech in Noise vs Time Compressed: $F(1, 3356) = 7.6, p = 0.005$

Vocoded vs Time Compressed: $F(1, 3356) = 2.36, p = 0.12$

Condition	Clear	Time Compressed	Vocoded	Speech in Noise
Mean	476.09	871.48	797.79	713.4
Std Dev.	399.15	452.7	507.34	411.47
Conf. Int.	[448.37, 503.81]	[838.05, 904.9]	[751.46, 844.13]	[678.87, 747.93]

Speaker:

Two vs Three: $F(1, 3356) = 3.15, p = 0.07$

Two vs Four: $F(1, 3356) = 0.14, p = 0.7$

Three vs Four: $F(1, 3356) = 4.64, p = 0.03$

Speaker	One	Two	Three	Four
Mean	560.6	674.23	733.27	839.17
Std Dev.	434.18	390.37	524.98	476.26
Conf. Int.	[526.02, 595.18]	[645.55, 702.91]	[692.64, 773.89]	[799.31, 879.02]

Block_trial_no *Condition:

Trial_no/Noise vs Trial_no/Vocoded: $F(1, 3356) = 2.04, p = 0.15$

Trial_no/Noise vs Trial_no/Compressed: $F(1, 3356) = 5.45, p = 0.01$

Trial_no/Vocoded vs Trial_no/Compressed: $F(1, 3356) = 0.49, p = 0.48$

Condition*Speaker:

Noise/Sp Two vs Noise/Sp Three: $F(1,3356) = 8.05, p = 0.004$

Noise/Sp Two vs Noise/Sp Four: $F(1,3356) = 0.44, p = 0.5$

Noise/Sp Three vs Noise/Sp Four: $F(1,3356) = 2.61, p = 0.1$

Vocoded/Sp Two vs Vocoded/Sp Three: $F(1,3356) = 4.28, p = 0.03$

Vocoded/Sp Two vs Vocoded/Sp Four: $F(1,3356) = 0.007, p = 0.92$

Vocoded/Sp Three vs Vocoded/Sp Four: $F(1,3356) = 6.36, p = 0.01$

Compressed/Sp Two vs Compressed/Sp Three: $F(1,3356) = 12.12, p < 0.001$

Compressed/Sp Two vs Compressed/Sp Four: $F(1,3356) = 3.65, p = 0.05$

Compressed/Sp Three vs Compressed/Sp Four: $F(1,3356) = 3.45, p = 0.06$

Appendix L

Lists of Speed and Capacity of Language Processing (SCOLP) sentences used in Experiment Seven to determine reading time.

REMOVED TO AVOID COPYRIGHT INFRINGEMENT

See the following sources for relevant information:

Baddeley, A., Emslie, H., & Nimmo-Smith, I. (1993). The Spot-the-Word test: a robust estimate of verbal intelligence based on lexical decision. *Br J Clin Psychol*, 32 (1), 55-65.

Baddley, A., Emslie, H., & Nimmo-Smith, I. (1992). *The Speech and Capacity of Language Processing Test manual*. Suffolk, UK: Thames Valley Test Company.

Appendix M

List of scrambled non-word sentences used the visual discrimination task of Experiment Seven.

Changed letters are highlighted in bold, capitalised font.

List A

Original Words	Non-words 1	Non-words 2
TEA HELPS PASS	AET PHE S L SSAP	AET PHE S L SSAP
COOKED BEFORE BELL	DCOEOK BE S PFF LBLE	DCOEOK BEROEF LBLE
CLAIM GOT RIGHT	LICAM OGT IRGTH	LICAM OGT IRGTH
HALF WAY HARDWEAR	HFAL YAW WR F BEAHR	HFAL YAW WREADAHR
RAN OUT SIXTEEN	ARN TUO TISEEXN	ARN TUO TISEEXN
SQUIRREL MAKES NICE	SLRQRIUE S L F NA INEC	SLRQRIUE SKEMA INEC
ENDED LATE JUNE	DDNEE ELAT JNEU	DDNEE ELAT JNEU
THISTLES BEND HIGH	E H M J UTST DBEN HHIG	EHLISTST DBEN HHIG
CASE NOW HARD	SECA NWO DAHR	SECA NWO DAHR
COLUMN PUT SUM	M UD P OL UPT MUS	MUCONL UPT MUS
SOAP TOP BATH	OSAP OTP BHTA	OSAP OTP BHTA
ADS SERVE CHEAT	DSA R T FFV HEACT	DSA RSEEV HEACT
THIS BEST BRAND	ITSH SEBT DARBN	ITSH SEBT DARBN
FAST WANT FINISH	TAFS NWTA SF J JIN	TAFS NWTA SFIIHN
WIT SAVED DAY	IWT SDVAE ADY	IWT SDVAE ADY
HARDSHIP KEEP HIM	HP T E I ARI EEPK IMH	HPSDHARI EEPK IMH
GENTLY WAKE HER	TELYGN EWAK EHR	TELYGN EWAK EHR
OATH OFFICE EACH	AOHT F I G FDO EHCA	AOHT FIFECO EHCA
VALVE RELEASE HOT	LVEVA LSEEAER HTO	LVEVA LSEEAER HTO
FREEZE WHEN COLD	RE F GFZ WNHE CODL	REEFEZ WNHE CODL
ASHES WORN OLD	ESAHS ONRW LDO	ESAHS ONRW LDO
PLAYED ON WHILE	EALYPD ON E X J IL	EALYPD ON EWIHL
HANDS WITH FRIENDLY	DAHSN THWI YELFDNIR	DAHSN THWI YELFDNIR
TRAY ACROSS GLASS	YART OSS D TA ASLGS	YART OSRCSA ASLGS
LEFT PARK SILVER	LTEF PRAK VSIREL	LTEF PRAK VSIREL
NEW GUESTS LEAVE	EWN SEGUTS E F B WL	EWN SEGUTS EEA V L
NO BILLS OFFICE	ON SLIBL CIEOFF	ON SLIBL CIEOFF
KITS DON'T FORGET	TKSI TOND TR H F P F	TKSI TOND TRGEOF
SAME TUNES EACH	ESAM SUENT EHCA	ESAM SUENT EHCA
CALLED NAME MANY	C L E F B L EANM MYNA	CLDEAL EANM MYNA

List B

Original Words	Non-words 1	Non-words 2
GREASE OFF DIRTY	E E S T BG FFO TDR I Y	EERSAG FFO TDR I Y

SQUARE STONE OVER	RAQUES TEOSN ROEV	RAQUES TEOSN ROEV
AIR PASSED THROUGH	IAR PDESSA TR <u>HI</u> UO	IAR PDESSA TRHGHUO
CRAWLED UNDER HIGH	WACERLD DNURE IGHH	WACERLD DNURE IGHH
PEAT AFTER LOGS	ATPE R <u>F</u> BGT SLOG	ATPE REAFT SLOG
PEACH PIE ICE	AHCEP IEP CEI	AHCEP IEP CEI
NEW SHOELACE THAT	NWE AL <u>P</u> DFESH ATHT	NWE ALOCEESH ATHT
TWO YOU ADD	OTW YUO DDA	OTW YUO DDA
ALMOST HURT SMALL	ST <u>N</u> M <u>P</u> A HUTR LMALS	STMLOA HUTR LMALS
THEY WERE NOISE	HTYE WEER NISOE	HTYE WEER NISOE
LAST NOVEL AT	SLTA <u>E</u> MP <u>W</u> N AT	SLTA ELOVN AT
CAME REAP OAT	MCAE PRAE TOA	MCAE PRAE TOA
DRIED WHEN MOVED	<u>D</u> F <u>J</u> S <u>D</u> EHWN ODMVE	DEIRD EHWN ODMVE
SURE ONE WAR	URSE ENO AWR	URSE ENO AWR
FOOD HOT CROSS	OFOD THO <u>S</u> D <u>S</u> P <u>S</u>	OFOD THO SCROS
YARD HAD MOLDY	RYAD HDA OYMLD	RYAD HDA OYMLD
STORE ROBBED LAST	ROSET ER <u>E</u> P <u>C</u> B LATS	ROSET ERDOBB LATS
LOT HELP FINISH	LTO LPEH FNSIHI	LTO LPEH FNSIHI
FIRM ON SHAKY	FMIR ON <u>Y</u> B <u>T</u> L <u>H</u>	FMIR ON YASKH
PULLED FRAIL CART	DPLLEU LIRFA CTAR	DPLLEU LIRFA CTAR
HOLD FOUR KINDS	HDOL FUOR <u>I</u> O <u>T</u> L <u>D</u>	HDOL FUOR INSKD
MUFF STYLISH ONCE	UFFM LITYSSH NCEO	UFFM LITYSSH NCEO
MAKES FINE BRAND	SKAME EIFN <u>B</u> E <u>O</u> S <u>A</u>	SKAME EIFN BDNRA
END ALL SUCH	DNE LAL CUHS	DNE LAL CUHS
THIN SHEET YELLOW	THNI <u>E</u> I <u>F</u> U <u>S</u> YOWLLE	THNI EHETS YOWLLE
EGGS TEA MUST	GEGS AET TSUM	GEGS AET TSUM
TINSEL FROM BOTH	IN <u>T</u> F <u>M</u> T OFRM BHOT	INSELT OFRM BHOT
CARD SLIP UNDER	CDRA SPLI DUREN	CDRA SPLI DUREN
DOUBT WAY WIND	<u>O</u> U <u>E</u> C <u>U</u> AYW NWID	OTDBU AYW NWID
HASH MADE RARE	HAHS AMDE ERRA	HAHS AMDE ERRA

List C

Original Words	Non-words 1	Non-words 2
WINDING PATH REACH	GNNIDIW HAPT HRECA	GNNIDIW HAPT HRECA
OVER FENCE PLUNGE	OVRE <u>F</u> O <u>F</u> C EUGLNP	OVRE FENEC EUGLNP
ROSE FROM EDGE	RSOE ROMF DEEG	RSOE ROMF DEEG
CLOSE BARN DOOR	<u>C</u> T <u>P</u> F <u>L</u> RNBA ORDO	CSOEL RNBA ORDO
CLOUD SEEN FARTHER	LUOCD NESE RRAFETH	LUOCD NESE RRAFETH
PLUS SEVEN LESS	LSPU <u>V</u> T <u>O</u> F <u>E</u> LSSE	LSPU VSNEE LSSE
HIGH LEVEL AIR	IGHH EELVL IAR	IGHH EELVL IAR
HISSED FROM BROKEN	<u>I</u> D <u>I</u> T <u>T</u> E MOFR KEOBRN	IDHSSE MOFR KEOBRN
STORES ACCOUNT LAST	ETOSRS CUAONT TSAL	ETOSRS CUAONT TSAL
HAVE COURT DECIDE	AEVH UORTC EC <u>J</u> E <u>E</u> E	AEVH UORTC ECIDDE

BROTHERS WEAR SAME	OERSBHTR RAEW ASEM	OERSBHTR RAEW ASEM
FORM OTHER NEED	RFOM <u>EIPUR</u> EEDN	RFOM EHOTR EEDN
DROPPED DOWN WORKMANS	PRDDEPO DNWO RNMKWSAO	PRDDEPO DNWO RNMKWSAO
HOIST UP TAKE	<u>IUITO</u> PU EKTA	ITHSO PU EKTA
SHEEP WHILE DOG	PEEHS WHIEL GDO	PEEHS WHIEL GDO
PAVED WITH STICKY	DPAEV TIWH <u>YSUDJK</u>	DPAEV TIWH YSTCIK
HEAT BRING OUT	TAHE BGNIR OUT	TAHE BGNIR OUT
MAKES FOR LACK	<u>MLBFS</u> OFR CLAK	MKAES OFR CLAK
OWED PAL THIRTY	EOWD APL RTTHIY	EOWD APL RTTHIY
GLASSES HELPED READ	LEGSSSA <u>PEFEIL</u> EDAR	LEGSSSA PEEDHL EDAR
WHEN SAY NASTY	WENH AYS ATSYN	WENH AYS ATSYN
TALES THEY TELL	<u>LFUBS</u> YHET ETLT	LETAS YHET ETLT
PRESSED FOR PAYMENT	RSPEEDS RFO TAMYPEN	RSPEEDS RFO TAMYPEN
FAIL HONEST MEN	IFAL <u>SOQUIE</u> ENM	IFAL SONTHE ENM
PAIL ONCE LET	IPAL OECN LTE	IPAL OECN LTE
MAIN BUSINESS NORTH	AMNI SIUNEBSS <u>HUPSN</u>	AMNI SIUNEBSS HTORN
BREW TEA COLD	BWRE TAE LCOD	BWRE TAE LCOD
YOUR BOOK FIRST	UYOR BKOO <u>SSUJF</u>	UYOR BKOO SRTIF
EVADE NET SWIM	EVDEA ENT WSIM	EVDEA ENT WSIM
MOSS GROWS NORTHERN	SOSM OGWSR <u>EOOSRTH</u>	SOSM OGWSR EONNRTH

List D

THERE WHEN SUN	<u>HFUFR</u> EWHN NUS	HETER EWHN NUS
LOAD YOUR LEFT	ADOL UYOR LTFE	ADOL UYOR LTFE
CLOSELY SIZE GAS	YL <u>FDMOS</u> SZIE AGS	YLECLOS SZIE AGS
GIRL LEFT DRUG	ILGR FETL UGDR	ILGR FETL UGDR
BEGAN RUST WHILE	GBAEN SRTU <u>WJIFL</u>	GBAEN SRTU WIHEL
YOUR PROBLEMS WISE	OURY MROSPBLE EISW	OURY MROSPBLE EISW
FLOOD MARK TEN	<u>FPMO</u> MAKR ETN	FDOLO MAKR ETN
BEST PART OF	TBES TPRA OF	TBES TPRA OF
HAD CROSS MEANT	DAH OSSRC <u>AUFNN</u>	DAH OSSRC ATEMN
SALT BEFORE FRY	ASTL ROEFEB YRF	ASTL ROEFEB YRF
LOG FLOAT WIDE	GLO <u>TBGMO</u> IDWE	GLO TAFLO IDWE
SWITCH CANNOT TURNED	SIWTCH COANTN RNUDET	SIWTCH COANTN RNUDET
CHAIR LEANED AGAINST	AICRH LN <u>FEFA</u> STAANGI	AICRH LNEDEA STAANGI
INK JAR STICKY	NKI RAJ IYTSKC	NKI RAJ IYTSKC
YOU DESIGNED FIT	OYU ES <u>HFJDDN</u> TIF	OYU ESGEIDDN TIF
RAIN MADE PLEASANT	RINA EMAD ESNPTAAL	RINA EMAD ESNPTAAL
CUTS TRIMS ANY	CTSU <u>SJNUR</u> YAN	CTSU SIMTR YAN

PAUSE BETWEEN THOUGHT	SPUAE BEETENW GTTHHOU	SPUAE BEETENW GTTHHOU
HORSE NOSE WAY	<u>EISPS</u> ENOS WAY	EHROS ENOS WAY
NECK MEANS NEAT	ECKN EANSM TNEA	ECKN EANSM TNEA
LANTERN WHICH GIVES	LANTENR HCWHI <u>EHWTI</u>	LANTENR HCWHI EGVSI
CROWDED WITH WILD	DEDOWCR HTIW LWDI	DEDOWCR HTIW LWDI
UNDER TENT SEE	<u>EESVN</u> ETTN EES	EDRUN ETTN EES
MUMBLE WILL SPEECH	MBMUEL IWLL ECESHP	MBMUEL IWLL ECESHP
DEEP SMELL PINEY	EDEP EMSLL <u>PFJOY</u>	EDEP EMSLL PEINY
DON'T ALWAYS SHOW	TNOD WYSAAL OSWH	TNOD WYSAAL OSWH
FRENCH WORSE THAN	EFRHCN <u>SXSFO</u> HNAT	EFRHCN SWREO HNAT
BILL BETWEEN TWO	ILLB EWBENTE WTO	ILLB EWBENTE WTO
ICE FROSTED PUNCH	CEI TRODESF <u>PVIOC</u>	CEI TRODESF PUHNC
ROSE SHOW WAS	ORSE WSOH WSA	ORSE WSOH WSA

Appendix N

Lists of Speed and Capacity of Language Processing (SCOLP) sentences used in Experiment Eight.

REMOVED TO AVOID COPYRIGHT INFRINGEMENT

See the following sources for relevant information:

Baddeley, A., Emslie, H., & Nimmo-Smith, I. (1993). The Spot-the-Word test: a robust estimate of verbal intelligence based on lexical decision. *Br J Clin Psychol*, 32 (1), 55-65.

Baddley, A., Emslie, H., & Nimmo-Smith, I. (1992). *The Speech and Capacity of Language Processing Test manual*. Suffolk, UK: Thames Valley Test Company.