MEASURING LISTENING EFFORT AND FATIGUE
IN ADULTS WITH HEARING IMPAIRMENT

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Biology Medicine and Health

2017

Sara W. Alhanbali

School of Health Sciences
Division of Human Communication, Development and Hearing
CONTENTS

TABLES ........................................................................................................................................... 5

FIGURES .......................................................................................................................................... 7

ABBREVIATIONS ............................................................................................................................ 9

ABSTRACT .......................................................................................................................................... 11

DECLARATION ................................................................................................................................... 12

COPYRIGHT STATEMENT ................................................................................................................ 12

DEDICATION ..................................................................................................................................... 14

ACKNOWLEDGMENT ...................................................................................................................... 15

CHAPTER ONE: INTRODUCTION .................................................................................................. 17

1.0 Background ............................................................................................................................... 17

1.1 Study one .................................................................................................................................... 22

1.2 Study Two ............................................................................................................................... 23

1.3 Study Three ............................................................................................................................ 25

1.4 Thesis format .......................................................................................................................... 26

CHAPTER TWO: BACKGROUND .................................................................................................. 29

2.0 Introduction ............................................................................................................................. 29

2.1 Listening effort and listening-related fatigue: definitions and models .................................. 30

2.1.1 Listening effort .................................................................................................................... 30
2.1.2 Listening-related fatigue ................................................................. 33

2.2 Quantifying self-reported listening effort and fatigue in individuals with hearing impairment ............................................................................................................. 35

2.3 Measures of listening effort and fatigue .............................................. 37

2.3.1 Measures of listening effort ................................................................. 37

2.3.2 Measures of listening-related fatigue ............................................... 62

2.4 The underlying dimensions assessed by self-report, behavioural, and physiological measures of listening effort and fatigue ........................................ 68

2.5 Summary and gap in knowledge ............................................................ 73

CHAPTER THREE: SELF-REPORTED LISTENING-RELATED EFFORT AND FATIGUE IN HEARING-IMPAIRED ADULTS ....................................................... 76

ADDENDUM TO STUDY ONE ....................................................................... 77

CHAPTER FOUR: HEARING HANDICAP AND SPEECH RECOGNITION CORRELATE WITH SELF REPORTED LISTENING EFFORT AND FATIGUE ....................................................... 79

ADDENDUM TO STUDY TWO ....................................................................... 80

CHAPTER FIVE: IS LISTENING EFFORT MULTIDIMENSIONAL? ................. 82

CHAPTER SIX: GENERAL DISCUSSION ....................................................... 137

CHAPTER SEVEN: CONCLUSIONS AND FUTURE DIRECTIONS ............... 144

7.0 Overview .............................................................................................. 144

7.1 Conclusions .......................................................................................... 144

7.2 Future directions ................................................................................... 146
TABLES

The page numbers of the tables in Chapters Three, Four, and Five, corresponds to the pagination of the manuscripts.

Chapter One: Introduction

Table 1. 1. Summary of each study in this PhD thesis including a description of research questions, participant details, outcome measures, listening tasks, statistical methods, findings, and corresponding documents. ................................................................. 21

Chapter Two: Background

Table 2.1. Examples of the different measures used in studies of listening effort............38
Table 2.2. Examples of dual tasks that have been used in published studies...............44

Chapter Three: Self-Reported Listening-Related Effort and Fatigue in Hearing-Impaired Adults (Study 1)

Table 1. Summary data for each group of participants.................................................e42
Table 2. FAS questions..................................................................................................e42
Table 3. EAS questions..................................................................................................e43
Table 4. Pattern matrix: Unique contribution by each variable to each factor..............e44

Chapter Four: Hearing Handicap and Speech Recognition Correlate with Self-Reported Listening Effort and Fatigue (Study 2)

Table 1. Results of regression analysis.........................................................................4
Chapter Five: Is Listening Effort is Multidimensional? (Study 3)

Table 1. The advantages and disadvantages of using self-report, behavioral, and physiological measures of listening effort ................................................................. 89

Table 2. Correlation coefficients between the test and the retest sessions of the different measures and results of ICC calculation with confidence intervals ....................... 112

Table 3. Matrix of the correlations between the measures ......................................................... 117

Table 4. Percentage of the variance explained by each factors .................................................... 118

Table 5. Pattern Matrix: factor loadings of the variables to each factor ................................. 119

Table 6. Correlation matrix between the factors ................................................................. 119
FIGURES

The page numbers of the figures in Chapters Three, Four, and Five, corresponds to the pagination of the manuscripts.

Chapter Two: Background

Figure 2.1. The new Ease of Language Understanding (ELU) model (Rönnberg et al., 2013) .......................................................... 33

Chapter Three: Self-Reported Listening-Related Effort and Fatigue in Hearing-Impaired Adults (Study 1)

Figure 1. Boxplots of fatigue assessment scale and effort assessment scale scores........e43

Figure 2. Scatter plot of fatigue assessment scale and effort assessment scale scores for all participants and the linear regression line.......................................................... e44

Chapter Four: Hearing Handicap and Speech Recognition Correlate with Self-Reported Listening Effort and Fatigue (Study 2)

Figure 1. Scatter plots showing age, Pure Tone Average (PTA), Signal to Noise Ratio (SNR), Hearing Handicap Inventory for Elderly (HHIE), Fatigue Assessment Scale (FAS), and Effort Assessment Scale (EAS) scores for all participants (n=84)......................................................... 3

Chapter Five: Is Listening Effort is Multidimensional (Study 3)

Figure 1. An outline of the sequence of events in each trial and the time periods used when analysing the data for each measure.................................................................102

Figure 2. Mean change in alpha power across participants and trials.................... 109
Figure 3. Correlations between the test (x-axis) and re-test (y-axis) data (n =30). RT: reaction time, SC: skin conductance. ................................................................. 113

Figure 4. Mean change in pupil size across participants and trials ........................................ 114
**ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANS</td>
<td>Autonomic Nervous System</td>
</tr>
<tr>
<td>ASA</td>
<td>Auditory Scene Analysis</td>
</tr>
<tr>
<td>CI</td>
<td>Cochlear Implant</td>
</tr>
<tr>
<td>CNS</td>
<td>Central Nervous System</td>
</tr>
<tr>
<td>dB HL</td>
<td>decibel Hearing level</td>
</tr>
<tr>
<td>EAS</td>
<td>Effort Assessment Scale</td>
</tr>
<tr>
<td>EEG</td>
<td>Electroencephalography</td>
</tr>
<tr>
<td>ELU</td>
<td>Ease of Language Understanding</td>
</tr>
<tr>
<td>EMG</td>
<td>Electromyography</td>
</tr>
<tr>
<td>ERBP</td>
<td>Event-Related Band Power</td>
</tr>
<tr>
<td>ERPs</td>
<td>Event Related Potentials</td>
</tr>
<tr>
<td>FA</td>
<td>Factor Analysis</td>
</tr>
<tr>
<td>FAS</td>
<td>Fatigue Assessment Scale</td>
</tr>
<tr>
<td>fMRI</td>
<td>functional Magnetic Resonance Imaging</td>
</tr>
<tr>
<td>FSS</td>
<td>Fatigue Severity Scale</td>
</tr>
<tr>
<td>FUEL</td>
<td>Framework for Understanding Effortful Listening</td>
</tr>
<tr>
<td>HA</td>
<td>Hearing Aid</td>
</tr>
<tr>
<td>HHIA</td>
<td>Hearing Handicap Inventory for Adults</td>
</tr>
<tr>
<td>HHIIE</td>
<td>Hearing Handicap Inventory for Elderly</td>
</tr>
<tr>
<td>HL</td>
<td>Hearing Loss</td>
</tr>
<tr>
<td>Hz</td>
<td>Hertz</td>
</tr>
<tr>
<td>ICC</td>
<td>Interclass Correlation Coefficient</td>
</tr>
</tbody>
</table>
IQR
KMO
MCM
MFI
MFSA-SF
MS
NASA TLX
PedSQL
POMS
PTA
RAMBPHO
SD
SNR
SSD
SSQ Hearing Scale
VAS-F
WM
µS

Interquartile Range
Kaiser-Meyer-Olkin
Motivational Control Model
Multidimensional Fatigue Inventory
Multi-dimensional Fatigue Symptom Inventory-Short Form
Multiple Sclerosis
NASA Task Load Index
Paediatric Quality of Life Inventory
Profile Of Mood States
Pure Tone Average
Rapid Automatic Multimodal Binding of Phonology
Standard Deviation
Signal to Noise Ratio
Single Sided Deafness
Speech, Spatial and Quality Hearing Scale
Visual analogue scale of Fatigue
Working Memory
Micro Siemens
ABSTRACT

Hearing loss increases the cognitive demands required to attend to, and understand, an auditory message. There are numerous anecdotal reports of sustained listening effort and fatigue in individuals with hearing loss. Therefore, listening effort and fatigue might be important consequences of hearing loss that are not captured by standard audiometric procedures.

The aim of the first study was to quantify real-world listening effort and fatigue in adults with hearing loss. Participants included 50 experienced hearing aid users, 50 cochlear implant users, 50 adults with single-sided deafness, and a control group of 50 adults with ‘good’ hearing. The study used the generic 10-item Fatigue Assessment Scale and a locally-developed 6-item Listening Effort Scale. The results revealed that all three groups of adults with hearing loss reported significantly greater listening effort and fatigue, relative to the control group. Listening effort (or fatigue) were not correlated with hearing level in the hearing aid group and there was no significant difference in mean effort/fatigue between the three groups.

The main aim of the second study was to investigate the correlation between hearing handicap and self-reported listening effort and fatigue. Participants included 86 adults with hearing loss, some of whom were hearing aid users. Handicap was measured using the 25-item Hearing Handicap Inventory for the Elderly. The results revealed a significant positive correlation between hearing handicap and both listening effort and fatigue. These findings are consistent with models and frameworks of listening effort and fatigue, which suggest that fatigue is a motivational control mechanism i.e., fatigue will be experienced if sustained effort is not perceived as rewarding.

During the preparation of this thesis, there has been an explosion of peer-reviewed publications on the topic of listening effort and fatigue; however, the literature is as confusing as it is voluminous: potential measures of listening effort and fatigue (self-report, behavioural, and physiological) frequently do not correlate with each other and sometimes result in contradictory findings. This raises questions about the sensitivity and reliability of the different measures along with the possibility that listening effort is a multidimensional phenomenon. Therefore, the aim of the final study was to investigate the reliability of potential measures of listening effort, to identify if they correlate with each other, and to use Factor Analysis to identify if the different measures tap into the same underlying dimension. Listening effort was measured simultaneously using multi-modal measures including: pupillometry, EEG alpha power, skin conductance, reaction time, and self-report. Recordings were obtained while 116 participants, with normal to severe hearing loss, performed a speech-in-noise task. Results revealed that the measures are mostly reliable. There were weak or non-significant correlations between the measures. Factor Analysis revealed that the measures grouped into four underlying dimensions, which we interpret as: i) performance, ii) cognitive processing, iii) alertness, and iv) behavioural consequences.

The findings of this PhD thesis revealed that high levels of listening effort and fatigue are common amongst adults with hearing loss. This suggests that a more comprehensive assessment of hearing disability should include measures of listening effort/fatigue. Further, the findings revealed that listening effort and fatigue correlate with perceived difficulties but not hearing level. The relationship between hearing level and effort/fatigue, like hearing impairment and hearing handicap, is not straightforward. Finally, measures of listening effort tap into independent dimensions. This latter finding provides a framework for understanding and interpreting listening effort, and has widespread implications for both research and clinical practice.
DECLARATION

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

COPYRIGHT STATEMENT

- The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

- Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

- The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.
Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property University IP Policy (see http://documents.manchester.ac.uk/display.aspx?DocID=24420), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.library.manchester.ac.uk/about/regulations/) and in the University’s policy on Presentation of Theses
DEDICATION

To Baba
I still remember the day I graduated from Kindergarten. You carried me as we went up the stairs because I was so tired to walk after the graduation party. I also remember your happy voice on the phone congratulating me for getting a high GPA in high school. I also remember your happy tears when I graduated as the top of my class in my Bachelors. Today, I’m about to finish my PhD and I don’t have you with me. It hurts me so much, I could not wish for anything more than having you by my side now. You are the reason I’m here today. You have always supported me to achieve what I want just because you believed it would make me happy.

Life might have taken you from me too soon. But you are always with me and in my heart. I wish to look in your eyes, hug you, and say thank you. My heart tells me that you feel me and that you have always been with me. Baba, you will always be my hero and the first prince in my life. May your soul rest in peace. I love you very much.
ACKNOWLEDGMENT

Getting to do this PhD was one of the greatest opportunities I had in my life. It has been and a very challenging yet extremely rewarding experience. I would have never been able to achieve anything without the endless love and support from my family, friends, and from the second family that I was so fortunate to have here in Manchester.

Kevin, I always have, and will always believe that you are a big part of everything I have achieved in these three years. I highly appreciate your belief in my capabilities to do a PhD. You gave me the motivation I needed to leave everything behind and come spend three years believing that I’m good enough to do a PhD. You brought the best in me and were there for me in the most difficult times. I learned so much from you and I will always feel so lucky that I got to be supervised by you. Million thank you(s) go to you. I am really glad that our paths crossed back in 2010!

Mohammad, my husband, my friend, and my endless love. You are my everything, I could not have done without your endless love, support, and understanding. I have always loved you because you are one clever man, but the help I got from you while doing this PhD made me realise how smart you actually are and made me love you even more! Missing you has also made learn how much I appreciate having you in my life.

My lovely family, Mama, Salah, and Abdulrahman. Thank you for always being there FOR me. Mama, you will always be my role model for a strong and independent woman.
Salah, my big brother and my backbone, you presence makes baba’s leaving less painful because I know that you always have my back. Abdulrahman my twin and my best friend, me and you know the big influence you had on pushing me do this PhD, highly appreciated brother!

My lovely friends and second family, Ghada, Reem, Hannah, Af, Chelsea, and Alex, you made Manchester feel like a second home for me. I would have never wished for better friends. Thank you for your endless care, support, and love. I love you guys so much.
Big thank you to my co-supervisor Piers Dawes who have always challenged me to do my best and helped me achieve things I have never thought I am capable of. My lovely advisors, Agnes Leger and Karolina Kluk-de kort, thanks a lot for helping me get through the toughest periods of my PhD. You both were more than great. Rebecca Millman, I consider myself very lucky to have you as a friend and as work collaborator. Thank you for everything you have done for me.

My gratitude also goes to every single ManCADer, such a great department and lovely people. Thank you for every single person who supported me, asked how I am, and tried to make me feel better when things became too much. I love you all very much.
CHAPTER ONE

INTRODUCTION
CHAPTER ONE

INTRODUCTION

1.0 Background

Individuals with hearing impairment commonly report the need to exert increased levels of listening effort in everyday listening situations despite using hearing aids or cochlear implants (Pichora-Fuller et al. 2016). Experiencing increased levels of listening effort for prolonged periods of time can result in the development of fatigue (McGarrigle et al. 2014). Increased listening effort and fatigue can have a negative impact on quality of life (Bess and Hornsby 2014). Despite being commonly reported by individuals with hearing impairment, self-reported listening effort and fatigue have not been quantified in individuals with hearing impairment. Establishing whether adults with hearing impairment do report higher levels of listening effort and fatigue compared to matching controls with good hearing is an important first step in justifying the relevance of the assessment of listening effort and fatigue within clinical audiology settings.

Measurement of listening effort and fatigue can improve our understanding of hearing disability by tapping into aspects of listening difficulties that are not readily assessed by standard audiometric measures. Establishing that hearing impairment is associated with listening effort and fatigue would justify the importance of identifying factors that contribute to these experiences and that can be targeted in hearing rehabilitation. For example, identifying how hearing sensitivity relates to self-reported listening effort and fatigue would inform hearing rehabilitation. A correlation between hearing sensitivity and both self-reported listening effort and fatigue would suggest that improving audibility
may reduce the experience of listening effort and fatigue. However, a lack of correlation between hearing sensitivity and both self-reported listening effort and fatigue would highlight the importance of identifying potential factors that are likely to influence individuals’ experience of listening effort and fatigue.

Self-report measures of listening effort and fatigue may assess an important part of hearing disability and provide insights to individuals’ perception of listening difficulties (McGarrigle et al. 2014). Objective measures may provide additional information about the physiological mechanisms underlying listening effort/fatigue. Objective measures may also provide a more accurate method for quantifying the benefits from hearing rehabilitation strategies than self-report measures (e.g. the benefit obtained from a hearing aid or a particular hearing aid signal processing algorithm on listening effort). There have been numerous attempts to measure listening effort and fatigue in research settings using self-report; e.g. Hornsby and Kipp (2016), behavioural; e.g. Houben et al. (2013), and physiological measures; e.g. Zekveld et al. (2010). Puzzlingly, although purporting to measure the same underlying dimension of “listening effort/fatigue”, self-report, behavioural, and physiological measures of listening effort/fatigue rarely correlate with each other and often result in contradictory findings across studies or across different groups of participants (McGarrigle et al. 2014). The contradictory findings raise questions about their reliability or/suggest the possibility that they may not tap into the same underlying dimension.

Establishing the reliability of potential measures of listening effort and fatigue is essential before they could be considered for use in clinical settings (Koo and Li 2016). Unreliable
measures are unlikely to correlate with each other. Therefore, establishing the reliability of the measures might help to eliminate a potential explanation for the absence of correlation between them. The variability in the testing methods used across different studies (e.g. experimental setup, test stimuli, participants, etc.) makes it difficult to identify whether measures of listening effort/fatigue tap into the same underlying dimension. An ideal method for identifying how different self-report, behavioural, and physiological measures relate to each other and whether they tap into the same underlying dimension would be to record the various alternative measures simultaneously during the same listening task. Identifying how measures of listening effort and fatigue relate to each other and whether they tap into the same underlying concept would provide a framework for understanding and interpreting the different measures and their underlying dimensions before they are considered for use in clinical practice.

Chapter Two of this thesis provides a background on the literature of listening effort and fatigue. It includes a discussion of the concepts of listening effort and fatigue, a discussion of the quantification of self-reported listening effort and fatigue within individuals with hearing impairment, a critical review of self-report, behavioural, and physiological measures of listening effort and fatigue, a discussion of the relationship between the different measures of listening effort/fatigue, and a discussion of the underlying dimensions assessed by the different measures. Due to the extensive nature of recent research on listening effort and fatigue, the focus of Chapter Two is on key research studies whose findings are relevant to the research questions addressed in this PhD thesis. Appendix A provides a list of references and summary of the findings of additional
recent research studies (published between 2015 and 2017) that were not mentioned in Chapter Two due to the word limit.

Table 1.1 provides a description of the main research questions addressed in each study of this PhD thesis in addition to a description of the methods used in each study including participant details, outcome measures, listening tasks, statistical methods, findings, and corresponding documents.
Table 1.1. Summary of each study in this PhD thesis including a description of research questions, participant details, outcome measures, listening tasks, statistical methods, findings, and corresponding documents. PTA: pure tone average at the frequencies 0.5, 1, 2, 4 kHz; HL: hearing loss; FAS: Fatigue Assessment Scale; EAS: Effort Assessment Scale; HHIE: Hearing Handicap Inventory for Elderly; NASA TLX: NASA Task Load Index; VAS-F: Visual Analogue Scale of Fatigue.

<table>
<thead>
<tr>
<th>Study</th>
<th>Research questions</th>
<th>Participant details</th>
<th>Outcome measures</th>
<th>Listening task</th>
<th>Statistical analysis</th>
<th>Main findings</th>
<th>Corresponding documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>• Do adults with hearing impairment report increased listening effort and fatigue compared to controls?  • Does effort correlate with fatigue and does effort/fatigue correlate with PTA?</td>
<td>n= 200 (55-85 years). PTA: normal to profound.</td>
<td>Self-report: FAS EAS</td>
<td>None.</td>
<td>• Kruskal-Wallis test and Mann-Whitney pair-wise test with Bonferroni correction.  • Spearman’s correlation coefficient.</td>
<td>• Greater listening effort/fatigue in adults with hearing impairment.  • Correlation between self-reported listening effort and fatigue.  • No correlation between hearing level and listening effort/fatigue.</td>
<td>Appendix B: FAS  Appendix C: EAS  Appendix D: PIS  Consent form</td>
</tr>
<tr>
<td>Two</td>
<td>• Does hearing handicap and speech recognition in noise correlate with self-reported listening effort/fatigue?</td>
<td>n= 84 (65-85 years). PTA: mild to severe.</td>
<td>Self-report: FAS, EAS, HHIE</td>
<td>Correct identification of digits triplets in noise.</td>
<td>• Spearman’s correlation coefficient.  • Multiple linear regression.</td>
<td>• Unlike hearing levels, hearing handicap (and to a lesser extent speech recognition) correlate with listening effort/fatigue.</td>
<td>Appendix F: HHIE  Appendix G: PIS  Consent form</td>
</tr>
<tr>
<td>Three</td>
<td>• Are potential measures of listening effort reliable?  • Do potential measures of listening effort correlate with each other and do they tap into the same underlying construct?</td>
<td>n= 116 (55-85 years). PTA: normal to severe.</td>
<td>Self-report: NASA TLX VAS-F  Behavioural: Reaction time Physiological: Pupillometry EEG Skin conductance</td>
<td>Recall correct digit from sequence of 6 presented in noise.</td>
<td>• Factor Analysis.  • Intra-class correlation coefficient with Spearman’s correlation.</td>
<td>• With the exception of skin conductance, measures have good reliability.  • Weak/absent correlations exist between the measures.  • Four dimensions underlie the measure: performance, alertness, processing, and behavioural consequences.</td>
<td>Appendix I: NASA TLX  Consent form</td>
</tr>
</tbody>
</table>

[Published in Ear and Hearing (Alhanbali et al. (2017a))]  
[Published in Ear and Hearing (Alhanbali et al. (2017b))]  
[Submitted for publication in Ear and Hearing]
1.1 Study one

The main aim of the first experimental study (Chapter Three) was to quantify self-reported listening effort and fatigue in adults with hearing impairment (hearing aid users, cochlear implants users, and individuals with single-sided deafness) and compare these with a control group of adults with good hearing. A secondary aim was to investigate the correlation between: i) self-reported listening effort and self-reported fatigue, and ii) hearing sensitivity and both listening effort and fatigue. Self-reported listening effort and fatigue were measured using two scales, the Effort Assessment Scale (EAS) and the Fatigue Assessment Scale [FAS; (Michielsen et al. 2004)], respectively.

Hypotheses:

1. Hearing-impaired participants report increased listening effort and fatigue compared to participants with good hearing. This hypothesis was based on the anecdotal reports of listening effort and fatigue from individuals with hearing impairment (McGarrigle et al. 2014).

2. Self-reported listening effort and fatigue correlate with hearing sensitivity. The hypothesis was based on the arguments that listening effort and fatigue are caused by the increased listening demands associated with the presence of hearing impairment (McGarrigle et al. 2014).

3. Self-reported listening effort correlates with self-reported fatigue. The hypothesis was based on the arguments that increased listening effort might result in the development of fatigue.

This study was published in *Ear and Hearing*: 

22

The Portable Document Format (PDF) of the reprint is used in the chapter.

### 1.2 Study Two

The purpose of the second experimental study (Chapter Four) was to identify correlates of self-reported listening effort and fatigue that can be targeted in hearing rehabilitation. The first aim was to investigate the correlation between hearing handicap (aka participation restrictions in the current International Classification of Functioning Disability and Health; World Health Organisation 2001) and both self-reported listening effort and fatigue. Hearing handicap was considered an indication of perceived communicative success. A second aim was to investigate the correlation between performance on a speech-in-noise-test and both self-reported listening effort and fatigue. Listening to speech in the presence of background noise is a more realistic task compared to the detection of pure-tones in quiet and thus might be more sensitive to listening effort and fatigue compared to hearing thresholds. The EAS and the FAS were used for the assessment of self-reported listening effort and fatigue, respectively. The Hearing Handicap Inventory for the Elderly [HHIE; (Ventry and Weinstein 1982)] was used as a measure of self-reported hearing handicap.
Hypotheses:

1. Self-reported hearing handicap correlates with both self-reported listening effort and fatigue. This hypothesis was based on the Motivation Control Model proposed by Hockey (2013). The Motivation Control Model suggests that fatigue is likely to be reported in cases of low motivation where sustained effort is perceived as not resulting in successful performance.

2. Self-reported listening effort and fatigue correlate with performance on a speech-in-noise task, but not so strongly as with self-reported hearing handicap. This is based on the fact that the speech-in-noise measure assesses some aspects of listening in everyday challenging listening situations, but is not truly representative of listening in such situations. The hypothesised strong correlation between the self-report measures used in the study (FAS, EAS, HHIE) was based on the fact that the three questionnaires assess difficulties that participants experience in everyday life.

The manuscript for this study has been accepted for publication in *Ear and Hearing*.


The PDF of the reprint is used in the chapter.
1.3 Study Three

The first aim of Study Three (Chapter Five) was to investigate the reliability of potential measures of listening effort and fatigue. Other aims included investigating the correlations between the measures and establishing whether they tap into similar or different underlying psychometric dimensions using Factor Analysis. Potential measures of listening effort and fatigue included: NASA Task Load Index, the Visual Analogue Scale of Fatigue, reaction time, pupillometry, skin conductance, and EEG alpha power.

Hypotheses:

1. Candidate measures of listening effort have good test-retest reliability. This is unknown and yet to be tested.

2. Candidate measures of listening effort do not correlate (or weakly correlate) with each other. This is based on previous research findings that did not report a correlation between potential measures of listening effort (e.g. Zekveld et al. 2010; Mackersie et al. 2015).

3. Candidate measures of listening effort load into a single common factor if they index the same construct.

At the time of writing, this manuscript has been submitted to *Ear and Hearing*.

Chapter One

The format used for submitting the manuscript to *Ear and Hearing* is used in the chapter.

### 1.4 Thesis format

The research carried out in this thesis has resulted in novel, publishable findings. Therefore, the “alternative format” used by the University of Manchester is appropriate for the presentation of the thesis. The alternative format also demonstrates the extent to which the candidate’s PhD training has cultivated skills in dissemination of research to readers of academic journals. The first author of each study is always the author of this thesis. For Chapters Three and Four, the PDF of the reprint is used. A page appears before each manuscript with the title of the manuscript and publication information. Chapter Three and Chapter Four will have their own pagination which does not follow the pagination of the rest of the thesis. Chapter Five (submitted to *Ear and Hearing*) will follow the pagination of the thesis. A page appears before Chapter Five with the title of the manuscript, details of the authors, and the journal name. A list of the references cited in each of the manuscripts is provided at its end. A list of all of the references cited in this thesis is provided at the end of the thesis. All references follow the referencing format of *Ear and Hearing*. The tables and figures in the thesis have also been formatted according to *Ear and Hearing* guidelines. The thesis combines both British English and American English, due to the submission of manuscripts to journals with American readerships. Hence, the manuscripts (Chapters Three, Four, and Five) contain American spellings, while the remainder of the thesis follows British spelling conventions.
For all of the experimental studies in this PhD thesis, co-authors, Kevin J. Munro and Piers Dawes suggested the main aim of the studies, advised on the design, analysis, results interpretation, and revised the manuscripts. The candidate has refined the research questions, designed the methods, conducted the data collection and analysis and drafted the manuscripts. Co-author Simon Lloyd contributed to the preparation of the first and the second manuscripts. Co-author Rebecca Millman contributed to the analysis and the interpretation of the results of the third experimental study and to the preparation of the manuscript.
CHAPTER TWO

BACKGROUND
CHAPTER TWO

BACKGROUND

2.0 Introduction

Experiencing increased levels of effort in challenging listening conditions is common among people with hearing impairment (Pichora-Fuller et al. 2016). Prolonged periods of effortful listening may result in fatigue (McGarrigle et al. 2014). Measurement of listening effort and fatigue is of interest in clinical audiology because these may tap into aspects of hearing disability that are not captured by standard clinical measures. Despite being commonly reported by individuals with hearing impairment, self-reported listening effort and fatigue have not been systematically quantified in individuals with hearing impairment yet this is an essential first step to justify the importance of their inclusion in a comprehensive assessment of hearing disability.

Various self-report, behavioural, and physiological measures have been used in the assessment of listening effort and fatigue in research settings (examples are provided in section 2.3). However, there is no consensus as to the most appropriate measure of listening effort or fatigue for research or clinical purposes, and the various measures do not always agree with each other (McGarrigle et al. 2014). Lack of agreement between measures raises questions about the reliability of the measures. Lack of agreement between the measures also suggests that listening effort and fatigue might be multidimensional phenomena with the different measures tapping into independent aspects of the same process. One of the limitations of purported measures of listening effort and fatigue is that some of them might be suitable for comparing groups of people
(i.e. for research purposes) but not for use on an individual basis in a clinical setting.

Some measures result in significant differences in listening effort between different conditions at the group level but not at the individual level (Dimitrijevic et al. 2017). In addition, some measures may be more suitable for testing certain populations; for example, measures that require dividing attention might not be ideal for testing children because of their limited ability to do so (Choi et al. 2008).

This review is divided into five sections: i) definitions of listening effort and fatigue and a discussion of models that explain these concepts; ii) quantification of self-reported listening effort and fatigue within individuals with hearing impairment; iii) self-report, behavioural, and physiological measures of listening effort and fatigue; iv) theories that propose the multidimensionality of listening effort; and v) summary and gap in knowledge.

2.1 Listening effort and listening-related fatigue: definitions and models

2.1.1 Listening effort

For individuals with normal hearing, listening is an automatic, effortless process in ideal listening conditions (McGarrigle et al. 2014). According to Mattys et al. (2012), degradation of auditory inputs can occur as a result of: i) factors related to the speaker, such as speaking in a non-native accent; ii) factors related to the environment, such as the presence of background noise; or iii) factors related to the listener, such as having a hearing loss. When perceiving degraded auditory inputs, people often report the need to “strain” or “work” to understand the auditory input (Hornsby and Kipp 2016).
Listening effort has been defined as “*the deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a task*” (Pichora-Fuller et al. 2016). Pichora-Fuller’s definition suggests that increased listening demands will not necessarily result in increased listening effort. However, the definition suggests that the deliberate allocation of cognitive resources is an essential aspect of listening effort. The motivation and the reward associated with task performance need to justify the exertion of increased effort. Effort associated with perceived successful task performance can be considered “*effective*”. This is likely rewarding and motivates further expenditure of cognitive resources. However, perceived failure to cope with the demands of the task despite of increased effort may be perceived as “*ineffective*” effort. This will decrease motivation to continue with the task. Further discussion on how motivation can affect task performance and individuals’ perception of listening effort is provided in sections 2.3 and 2.4.

In order to improve the understanding of the concept of listening effort, Rönnberg and colleagues have provided an explanation of listening effort using The Ease of Language Understanding (ELU) model (Rönnberg et al. 2008; Rönnberg et al. 2013). According to the ELU model, listening effort occurs when automatic (bottom-up) speech processing is interrupted. The model suggests that the working memory (WM) has an essential role in language comprehension. WM is defined as “*a limited capacity system for temporarily storing and processing the information required to carry out complex cognitive tasks such as comprehension, learning, and reasoning*” (Rönnberg et al. 2013). WM has been integrated into speech perception models because it is responsible for being able to
perform tasks that include processing and storing of elements, two components that are crucial for speech understanding (Rudner et al. 2012).

According to the ELU, the linguistic content of inputs (phonology, semantics, syntax, and prosody) received via any language perception mode are combined into what the authors refer to as Rapid Automatic Multimodal Binding of Phonology (RAMBPHO). RAMBPHO is then automatically compared with similar information in the long-term memory. In the case of a match between the RAMBPHO and information in the long-term memory, the message will be automatically understood. This has been described as an automatic process, referred to as “implicit processing”. Suboptimal listening conditions result in degradation of the signal and consequent failure to identify a match between the signal and information in the long-term memory. In the case of mismatch, “explicit processing” takes place and the involvement of the WM is required to understand the degraded message. The WM resolves the ambiguity of the signal by using the context to “fill in the gaps” of the perceived signal based on previous knowledge and experience. Figure 2.1 provides an outline of the explicit and implicit speech processing described in the ELU. In explicit processing, cognitive processes are required in order to “untangle” the input to obtain a match with information in the long-term memory (Picou et al. 2011). According to the ELU model, increased cognitive processing in challenging listening conditions is the basis for the concept of “listening effort”.

Edwards (2016) has recently elaborated on the ELU model by incorporating the concept of Auditory Scene Analysis (ASA) into the model (hybrid ASA and ELU model). Edwards defined ASA as “the organisation of auditory signal components into perceptually
meaningful objects” (Edwards 2016). According to Edwards, increased cognitive load can occur during ASA, i.e. during the synthesis of inputs that will later be compared with information in the long-term memory. This initial step is not considered in the ELU model. Edward’s hybrid ASA and ELU model suggests that the quality of the input signal can have a great impact on the cognitive load during ASA. The more degraded the input, the greater the cognitive load required before the stage of implicit or explicit processing described in the ELU model takes place. The elaboration of the model suggested by Edwards also applies to the perception of environmental sounds. Individuals with hearing impairment commonly report poor awareness of environmental sounds. For example, an individual with hearing impairment might find it difficult to identify that someone is walking behind him even when the footsteps are audible.

Figure 2.1. The new Ease of Language Understanding (ELU) model (Rönnberg et al., 2013).

2.1.2 Listening-related fatigue

Fatigue is a common experience in individuals with chronic health conditions. Some health conditions, e.g., multiple sclerosis, result in physical fatigue (Krupp and Christodoulou 2001). Other health conditions, e.g., traumatic brain injury, result in mental fatigue as a result of compromised cognitive processing capacity (Belmont et al. 2006). Some health conditions can also result in emotional fatigue, which is characterised by lack of motivation in to engage in physical or mental tasks due to increased psychological
demands (Hornsby et al. 2016); an example is the effect that cancer treatment can have on cancer patients. The distinction between different types of fatigue is not always clear as they can sometimes influence each other (Hornsby et al. 2016). For example, a student might feel lacking energy to go out after a long day of lectures. Fatigue may adversely affect patients’ quality of life (Hornsby and Kipp 2016). Fatigue can result in decreased productivity and increased chance of work-related injuries (Ricci et al. 2007). Fatigue can also have negative psychological impacts such as depression and lack of desire to engage in daily life activities and social interactions (Ferrando et al. 1998).

Fatigue may develop in individuals with hearing impairment as a result of the increased cognitive demands they experience in everyday listening situations (Bess and Hornsby 2014). According to the research of Bess and Hornsby, individuals who report listening effort on a daily basis also report feeling exhausted, tired, and lacking energy at the end of the day. Fatigue has been found to have a negative impact on the lives of individuals with hearing impairment. For example, Kramer et al. (2006) reported that workers who had hearing impairment required significantly more sick leave than matching controls with normal hearing. Nachtegaal et al. (2009) have also reported that participants with hearing impairment require more time to recover from work because of the increased listening effort they experience. The negative impact of sustained effortful listening gave rise to the concept of “listening-related fatigue” which has been defined as “extreme tiredness resulting from effortful listening” (McGarrigle et al. 2014).

Traditionally, listening related fatigue has been considered a direct consequence of effortful listening. However, a number of models and frameworks suggest that fatigue is
likely to develop in demanding listening situations that are perceived as being unrewarding (Hockey 2013; Pichora-Fuller et al. 2016). According the Motivational Control Model (MCM) developed by Hockey (2013), motivation to engage in task performance is unlikely to result in the development of fatigue as long as increased effort is reinforced by perceived successful performance. The hypothesised association between effort, fatigue, perceived performance, and motivation has important implications for clinical and research purposes. The hypothesised associations highlight the importance of identifying correlates of listening effort and fatigue that can be targeted in hearing rehabilitation. A positive correlation between perceived hearing difficulty and both listening effort and fatigue would suggest that a hearing rehabilitation process with a focus on psychological factors such as motivation may improve outcome. More details on the association between perceived hearing difficulty and self-reported listening effort and fatigue are provided in Chapter Four. The MCM suggests that behavioural and physiological measures of listening effort/fatigue might not always correlate with self-report measures. For example, listening effort/fatigue measured using behavioural or physiological measures would not necessarily translate into a perceived state of listening effort/fatigue in cases of high motivation or when increased effort is rewarded by perceived successful performance.

2.2 Quantifying self-reported listening effort and fatigue in individuals with hearing impairment

Fatigue was found to have negative impacts on the quality of life of patients with chronic health conditions such as cancer (Stone et al. 2000), and Parkinson’s disease (Brown et al. 2005). In these health conditions, the prevalence of fatigue was found to be relatively
high and it has therefore been routinely measured in all patients. Generic and disease-specific fatigue scales have been also developed for the assessment of fatigue in patients with chronic health conditions. Unlike fatigue, the assessment of effort has not received much attention in the daily life of patients with chronic health conditions probably because of its transient nature.

Despite the negative impact that listening effort and fatigue can have on the quality of life of individuals with hearing impairment, the assessment of listening effort and fatigue in daily life has not received much attention. Reports of increased listening effort and fatigue are mostly anecdotal. Establishing whether individuals with hearing impairment report increased listening effort and fatigue compared to controls with good hearing is an essential first step to justify the importance of their assessment within a clinical setting. Current self-report measures of hearing disability do not usually include items about listening effort or fatigue. One exception is the effort related items in the Speech, Spatial, and Quality (SSQ) Hearing Scale (Gatehouse and Noble 2004). However, the SSQ Hearing Scale is not routinely used in clinical settings. The absence of a hearing-specific scale of listening effort or fatigue justifies the use of generic scales used in the assessment of other chronic health conditions (further details are provided in Chapter Three). During the time this PhD was completed, one study investigated self-reported fatigue and vigour in adults with hearing impairment using generic fatigue scales (Hornsby and Kipp 2016). The authors reported decreased self-reported vigour in adults with hearing impairment compared to a matching control group (see section 2.3.2.1 for further details).
2.3 Measures of listening effort and fatigue

The inclusion of measures of listening effort and fatigue as dimensions of hearing disability may be of value in informing effective interventions. Measures of listening effort and fatigue might improve hearing rehabilitation by highlighting aspects of listening difficulties that are not readily assessed by standard audiometric measures. Understanding which aspects of hearing disability are particularly problematic for individual patients may also assist in making decisions about treatment and management options; for example, when the provision of a hearing aid is questionable in terms of the need to restore audibility, as in the case of mild hearing loss (McGarrigle et al., 2014). The following sections provide an overview of the self-report, behavioural, and physiological measures that have been used in the assessment of listening effort and fatigue. Appendix A also provides a summary of a number recent studies (published between 2015 and 2017) that have used behavioural or physiological measures in the assessment of listening effort but were not mentioned in the following sections due to the word limits.

2.3.1 Measures of listening effort

A number of self-report, behavioural, and physiological measures have been used in the assessment of listening effort in research settings. Table 2.1 provides examples of the different measures of listening effort.
Table 2.1. Examples of the different measures used in studies of listening effort.

<table>
<thead>
<tr>
<th>Measures of listening effort</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-report measures</td>
<td>Effort-related questions in the SSQ Hearing Scale (Gatehouse and Noble 2004).</td>
</tr>
<tr>
<td>Behavioural measures (single task)</td>
<td>Response time to verbal, visual, or vibrotactile stimuli, e.g., Houben et al. (2013).</td>
</tr>
<tr>
<td>Behavioural measures (dual task)</td>
<td>Sentence repetition as the primary task and visual tracking as the secondary task, e.g., Desjardins and Doherty (2013).</td>
</tr>
<tr>
<td>Physiological measures (changes in the central nervous system)</td>
<td>Electroencephalography, e.g., Obleser and Kotz (2011)</td>
</tr>
<tr>
<td>Physiological measures (changes in the autonomic nervous system)</td>
<td>Skin conductance, e.g., Mackersie et al. (2015); pupil dilation, e.g., Zekveld et al. (2011).</td>
</tr>
</tbody>
</table>

2.3.1.1 Self-report measures of listening effort

Self-report measures rely on the patients’ reported experience of listening effort. The value of self-report measures should not be underestimated, since behavioural or physiological indications of listening effort/fatigue would be of minimal practical importance if listening effort/fatigue was not subjectively reported. Self-reported listening effort in everyday life can be measured using the effort-related questions in the SSQ Hearing Scale: i) Do you have to put in a lot of effort to hear what is being said in conversation with others?; ii) How much do you have to concentrate when listening to someone?; iii) How easily can you ignore other sounds when trying to listen to something? Listening effort resulting from performing a listening task can be measured using scales designed to assess the demands associated with task performance such as the NASA Task Load Index (Hart and Staveland 1988). Self-report measures have the advantage of being quick, easy, and inexpensive to administer (Bess and Hornsby 2014).

Self-report measures of listening effort have not been used previously to assess listening effort in the everyday life of adults with hearing impairment. However, self-report
measures have been frequently used in research settings to: i) investigate self-reported listening effort during the performance of a demanding listening task, e.g. Zekveld et al. (2010); ii) assess hearing aid benefit, e.g. Hornsby (2013); and iii) compare listening effort in different groups of participants, such as participants with normal hearing and participants with hearing impairment, e.g. Mackersie et al. (2015).

Puzzlingly, self-report measures of listening effort do not typically correlate with other behavioural/physiological measures of listening effort. Zekveld and colleagues (2011) reported no correlation between changes in pupil size and self-reported listening effort when presenting participants with sentences in different levels of background noise. The authors suggested that pupillometry and self-reported measures assess independent aspects of listening effort. The findings of Zekveld and colleagues suggest that pupillometry and self-report measures cannot be used interchangeably. The findings also suggest that multiple dimensions might need to be considered in the assessment of listening effort.

Self-report measures may not always be sensitive to the increased listening demands imposed on individuals with hearing impairment in research settings. For example, Mackersie et al. (2015) did not identify a difference in self-reported listening effort between participants with normal hearing and participants with hearing impairment. The authors suggested that the lack of a difference between the groups might be because the group of participants with hearing impairment had previous experience of performing listening tasks in research setting in contrast to participants with normal hearing who performed a research experiment for the first time.
Other limitations may apply to self-report measures of listening effort. Firstly, individuals might have different standards for judging how “effortful” a task is and this may influence their subjective ratings of effort in research settings. Having a hearing problem is likely to cause individuals with hearing impairment to experience increased listening demands more frequently in everyday life than individuals with normal hearing. In challenging experiments, a participant with normal hearing might report an equal experience of listening effort to a participant with hearing impairment because of lack of experience with challenging listening conditions. Secondly, self-report measures may be affected by the way “effort” is interpreted by participants. For example, some participants might rate their performance on the task rather than the effort they exerted (McGarrigle et al. 2014). Thirdly, self-report measures of effort do not explain the physiologic process underlying “effort” (Bess and Hornsby 2014). Identifying the underlying physiological mechanisms of listening effort would help in understanding its dimensionality and in identifying whether multiple measures are required for its assessment. The limitations associated with the use of self-report measures might have contributed to the lack of sensitivity of self-report measures to the hypothesised increased listening effort in some research studies, e.g. Mackersie et al. (2015). The influence of the aforementioned limitations on the subjective reports of listening effort might have contributed to the lack of correlation that has often been reported between self-report measures and other behavioural/physiological measures of listening effort e.g. (Hornsby 2013; Mackersie et al. 2015).
2.3.1.2 Behavioural measures of listening effort

Behavioural measures assess the effect of increased listening effort on aspects of task performance such as accuracy and speed of processing. Unlike audiological measures of word recognition, behavioural measures of listening effort may indicate an increased listening demand before task difficulty affects the accuracy of performance, e.g. the speed of processing can slow down without being associated with incorrect responses. The effect of increased listening demands on performance is an indirect indication of cognitive effort (Bess and Hornsby 2014). Both single- and dual-task paradigms have been used for measuring listening effort, as outlined below.

Single task

Single-task paradigms based on response times to verbal inputs have previously been used by several independent researchers. Response time has been used to index the impact of an assistive listening device (a hearing aid), e.g. Gatehouse and Gordon (1990); to assess listening effort in unfavourable listening conditions such as the presence of background noise, e.g. Picou et al. (2011); and to assess listening effort associated with performing a cognitively demanding task such as mental calculations on digits presented in background noise, e.g. Houben et al. (2013). When using response-time paradigms, it is argued that challenging listening tasks result in increased response times as a result of increased listening effort.

Gatehouse and Gordon (1990) were the first to use a behavioural measure of listening effort and did so for the purpose of evaluating the benefit obtained from hearing aid use. Response times were measured in aided and unaided conditions for several tasks:
detection of pure tones and speech shaped noise and recognition of single words and sentences. For the recognition tasks, response times were measured only for correctly identified words and sentences. Response times decreased when participants performed the listening task with their hearing aids on. A greater difference in response time between the aided and the unaided condition was identified when the material presented was speech. The findings of Gatehouse and Gordon’s study are consistent with the hypothesis that hearing aids result in increased benefit at the perceptual level, i.e. when speech understanding is required. The findings of Gatehouse and Gordon suggest that reaction time provided information about the listening demands imposed on the participants when increased task demands did not have a negative effect on the performance accuracy of the conventional measure, i.e. correct word.

Houben et al. (2013) used response time to measure listening effort in participants with normal hearing. Participants were presented with digit triplets from the Dutch Digit Triplets Test (Smits et al. 2004). Digits were presented in quiet and at high and low signal-to-noise ratios (SNRs). In the first task, participants had to identify the last digit in the triplet (i.e. an identification task). In the subsequent task, participants had to calculate and report the sum of the first and the last digits in the triplet (i.e. an additional arithmetic task). The arithmetic task was used in an attempt to increase the cognitive demands associated with task performance and approximate speech processing demands in everyday listening situations where cognitive processing of information is required. There was a significant main effect of SNR on the performance of both tasks, with slower response times in the more adverse SNRs. There was also a significant main effect of task on response time across the different SNRs. Response times were longer in the arithmetic
task compared to the identification task. Therefore, the findings of Houben and colleagues (2013) might suggest that a task requiring both retention and processing of information, such as performing mental calculations, might be more cognitively demanding and thus more sensitive to changes in listening effort than tasks requiring repetition of inputs. Houben and colleagues suggested that response time measures to simple speech stimuli (such as the identification and the arithmetic tasks) may provide an informative index of listening effort.

One of the limitations of using response time as a measure of listening effort is that it is not a “process-pure” measure (Pichora-Fuller et al. 2016). An assumption of the response-time paradigm is that increased task difficulty results in longer processing time, as more cognitive “work” is required to recognise and respond to stimuli. However, increased processing time might not necessarily be perceived as more effort. Increased task difficulty could result in increased expenditure of effort to maintain the same level of performance, with no difference in response time despite of increased effort (Bess and Hornsby 2014). Increased effort to maintain task performance may also result in shorter response time. Further research is required to identify how the underlying dimension of increased listening demands assessed by response time relates to other physiological and self-report measures of listening effort. More research is also required to identify the sensitivity of response time to the increased listening demands imposed on individuals with hearing impairment.
**Dual tasks**

Dual tasks for measuring listening effort are based on the “limited capacity” model developed by Kahneman (1973). Kahneman’s model proposes that individuals have finite cognitive resources available for task performance. When performing two or more tasks simultaneously, most cognitive resources will be directed towards performing the primary task. Any spare capacity would be utilised in performing the secondary task. As the primary task becomes more demanding, more cognitive recourses will be directed towards its performance leading to a deterioration in the performance of the secondary task (Downs and Crum 1978).

When using the dual task as a measure of listening effort, the primary task is a listening task, e.g. words or sentences presented in different levels, or types, of background noise. Secondary tasks have sometimes involved the auditory domain, e.g. tone detection (Hicks and Tharpe 2002), the visual domain, e.g. detecting a flash light (Sarampalis et al. 2009), or responding to a vibrotactile stimulus (Fraser et al. 2010). Table 2.2 provides examples of different dual tasks.

<table>
<thead>
<tr>
<th>Study</th>
<th>Primary task</th>
<th>Secondary task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fraser et al. (2010)</td>
<td>Sentence recognition</td>
<td>Tactile pattern recognition</td>
</tr>
<tr>
<td>Howard et al. (2010)</td>
<td>Repetition of monosyllabic words</td>
<td>Serial recall (rehearsal of a series of digits for later recall)</td>
</tr>
<tr>
<td>Sarampalis et al. (2009)</td>
<td>Sentence repetition</td>
<td>Response time to visual stimuli</td>
</tr>
<tr>
<td>Desjardins and Doherty (2013)</td>
<td>Repetition of sentences presented in three different types of maskers</td>
<td>Visual tracking</td>
</tr>
</tbody>
</table>
One of the advantages of the dual task paradigm is that it simulates some real-world listening situations where multi-tasking is required. For example, in a classroom environment, students are required to listen, understand, and take notes at the same time (McGarrigle et al. 2014). Some studies reported decrements in performance on a secondary task among individuals with hearing impairment (Downs 1982) or in older versus younger adults (Tun et al. 2009; Gosselin and Gagne 2011). However, the findings of studies that used dual-task measures of listening effort have not always been consistent. Improved secondary-task performance has sometimes been reported as a result of using hearing aids or noise reduction algorithms, e.g. Sarampalis et al. (2009) and Hornsby (2013). However, on other occasions, using noise reduction algorithms improved secondary-task performance only in participants with limited WM capacity (Neher et al. 2014). Both Hicks and Tharpe (2002) and Howard et al. (2010) reported deterioration in the performance of a secondary task in children with hearing impairment compared to children with normal hearing, interpreting this finding as an indication of increased listening effort. On the other hand, Desjardins and Doherty (2013) did not identify an effect of hearing impairment on the performance of a secondary task in older adults with hearing impairment compared to older adults with normal hearing. Different primary and secondary tasks may tax the cognitive system in different ways and to different extents, making it difficult to compare results (Ohlenforst et al. 2017a).

Performance on the dual task often does not correlate with participants’ self-reported listening effort. For example, Gosselin and Gane (2011) reported no correlation between secondary-task performance and self-reported listening effort when groups of older and younger adults preformed a dual task that involved a listening task (primary task) and a
tactile pattern recognition task (secondary task). In addition, contradictory findings were sometimes obtained when the dual task was used along with a self-report measure of listening effort. Hornsby (2013) found that the use of a hearing aid resulted in improved performance on a dual-task paradigm that involved a word recognition task (primary task) and a visual response time task (secondary task). However, there was no difference in self-reported listening effort between the aided and the unaided conditions.

The lack of agreement between the dual task and self-report measures suggests that the two measures may tap into independent aspects of listening effort. Dual tasks assess participants’ multi-tasking abilities but may not be related to the concept of “listening effort”. The link between listening effort and multitasking relies on certain assumptions that might not be entirely correct (McGarrigle et al. 2014). An assumption of the dual task paradigm is that the entire cognitive capacity will be utilised in performing the primary and the secondary tasks. However, it is not possible to identify whether participants do use their entire cognitive capacity or not. Additionally, the assumption that people always prioritise performance of the primary task is questionable. Participants might direct their attention towards the easier task regardless of instructions to prioritise the primary task (Styles 2006).

2.3.1.3 Physiological measures of listening effort

Physiological measures of listening effort include measures of changes in the central nervous system (CNS) and measures of changes in the autonomic nervous system (ANS). Measures of changes in the CNS include: i) functional Magnetic Resonance Imaging
(fMRI), e.g. Wild et al. (2012); ii) electroencephalography (EEG), e.g. Petersen et al. (2015); and iii) event related potentials (ERPs), e.g. Obleser and Kotz (2011). Measures of changes in the ANS include: i) pupil dilation, e.g. Zekveld et al. (2010); ii) skin conductance; iii) skin temperature; iv) heart rate; v) electromyographic activity, e.g. Mackersie and Cones (2011) and Mackersie et al. (2015); and vi) pre-ejection period, which is the time period between the excitation of the left ventricle in the heart and the opening of the aortic valve, e.g. Richter (2016).

**Functional Magnetic Resonance Imaging (fMRI)**

fMRI can be used to assess the changes in blood flow in the brain that occur as a result of increased levels of attention associated with listening effort (Wild et al. 2012). Wild and colleagues presented participants with speech stimuli degraded using a noise-vocoding technique to create a total of 4 different conditions (including clear speech). Vocoded sentences were created by filtering the speech signal to create a range of stimuli with different levels of difficulty. In each trial, participants were presented simultaneously with a speech stimulus, an auditory distracter (a series of tone bursts separated by variable durations of silence), and a visual distracter (series of cross-hatched white ellipses on a black background), and were instructed to attend to one of the three. When participants were instructed to attend to speech inputs, increased blood flow was measured in the frontal cortical regions (including left inferior frontal gyrus). Increased blood flow was identified when presenting degraded speech compared to clear speech and as the degradation of the signal increased. Increased blood flow in the frontal cortex was interpreted as an indication of increased attention and a neural marker of listening effort.
The authors concluded that attention (indicated by enhanced blood flow to the frontal cortex) was a crucial requirement for processing the degraded speech signal.

fMRI provides information about the location of brain activity associated with task performance. However, the use of fMRI in listening effort research has some limitations. fMRI produces acoustic noise that can affect the perception of the auditory stimulus (Blackman and Hall 2011). fMRI also has low temporal resolution which limits its ability to detect rapid task-evoked responses (Obleser et al. 2012).

**Electroencephalography (EEG)**

Changes in EEG activity can be classified as induced and evoked activity (Cohen 2014). The main difference between induced and evoked EEG activity relates to the time of the response. Induced EEG activity occurs at any point following the presentation of a stimulus while evoked EEG activity is characterised by a response that is time locked to the stimulus (Cohen 2014).

**Induced EEG activity**

In listening-effort research, changes in alpha (8-12 Hz) and theta (4-8 Hz) band activity have been used as indicators of listening effort, e.g. Obleser and Weisz (2012) and Wisniewski et al. (2015). According to inhibition theory, synchronised alpha activity is likely to occur when trying to memorise learned information during a stimulus free period (Klimesch et al. 2007). Increased alpha activity is also an indication of suppression of irrelevant stimuli such as suppression of background noise during effortful listening (Sauseng and Klimesch 2008). EEG has been used to assess the effect of listening to
degraded speech inputs on alpha activity, e.g. Obleser and Kotz (2011), and on theta activity, e.g. Wisniewski et al. (2015). EEG has been also used to assess the effect of hearing impairment on listening effort, e.g. Petersen et al. (2015).

Obleser and Kotz (2011) had participants rate the comprehensibility of mono-, bi-, and tri-syllabic nouns degraded using a noise-vocoding technique. Alpha activity was recorded while participants listened to the speech material. Increased alpha power during speech presentation was proportionate to the amount of acoustic degradation of the signal with increased acoustic degradation yielding increased alpha power. Obleser and Kotz also reported increased alpha power during a stimulus-free period that followed the presentation of speech where participants were asked to memorise what they heard (retention period). Increased alpha power during the retention period was also proportionate to the degradation of the speech signal. Consistent with inhibition theory (Klimesch et al. 2007), the authors suggested that increased alpha power during the retention period is an indication of increased demands on WM to retain learned information. Increased alpha power during a retention period has also been reported in a number of other studies, e.g. Obleser et al (2012).

Increased alpha power in adverse listening conditions has been considered an indication of increased listening effort (Obleser and Kotz 2011; Obleser and Weisz 2012). However, results have not always been consistent. Based on the findings of Obleser and Kotz (2011) and Obleser et al. (2012), Petersen et al. (2015) hypothesised that increased listening demands (memory demands, amount of acoustic degradation of the signal, and the presence of hearing impairment) should result in increased alpha power. There were two
groups of participants; one with good hearing and one with mild to moderate hearing loss. The authors also hypothesised that increased alpha power would correlate with hearing sensitivity. Participants were presented with digits that they were instructed to memorise during a retention period. The level of background noise and the memory demands varied across conditions. In contrast to the findings of Obleser and Kotz (2011), no difference in alpha power was identified between the listening conditions during the speech presentation period. In addition, no difference in alpha power was identified between the two groups of participants during the speech presentation period. During the retention period, alpha power increased as memory demands and signal degradation increased in participants with normal hearing and with hearing impairment. However, in the most demanding listening condition, alpha power only increased in participants with normal hearing or mild hearing loss. Alpha power decreased significantly in participants with moderate hearing loss. The authors suggested that alpha power might have reached a ceiling in participants with moderate hearing loss after which no further increase can occur as a result of limited cognitive resources. The authors suggested that decreased alpha power is an indication that participants with moderate hearing loss might have run out of cognitive resources. Participants with moderate hearing loss have likely exerted most cognitive resources when trying to identify the digits and thus had least spare cognitive resources available for memorising the digits. The explanation provided by the authors does not account for the absence of any effect of cognitive-resource depletion on the accuracy of task performance. Decreased alpha power may also occur as a result of losing motivation when performing a challenging listening task. However, losing motivation would likely result in impaired task performance. Future work should consider investigating factors that might contribute to suppressed alpha power in challenging
conditions by varying aspects as participants’ motivation when performing challenging listening tasks.

McMahon et al. (2016) recorded EEG alpha power and changes in pupil size in participants with normal hearing while listening to vocoded sentences in the presence of background noise. The authors reported increased alpha power compared to baseline during speech presentation. Their findings are consistent with the inhibition theory, whereby increased alpha power is an indication of suppression of background noise. In contrast to the findings of Obleser and Kotz (2011), alpha power was greater when participants attended to the more-intelligible 12-channel vocoded sentences than when they attended to the less-intelligible 6-channel vocoded sentences. The authors suggested that using a different listening task to the one used in Obleser and Kotz’s study might have contributed to the inconsistent findings because different tasks might tax the cognitive system in different ways. The difference in alpha power between the 6- and the 12-channel vocoded sentences was only identified at the most challenging SNR (≤ -3 dB). Changes in alpha power did not correlate with changes in pupil size. Differences in pupil size between the listening conditions were only identified at the least challenging SNRs (≥ 1 dB). The authors suggested that the EEG alpha power and pupillometry might be driven by different neurophysiological or attentional mechanisms. The lack of agreement between pupillometry and EEG alpha power highlights the importance of investigating the underlying dimensions assessed by each of these measures. Currently, researchers use both of the measures interchangeably to index listening effort. However, the finding that they show a different pattern of change across different levels of background noise suggests that each of them might index an independent aspect of listening effort.
Changes in pupil size perhaps index arousal levels while changes in alpha power might relate to cognitive processing associated with speech perception. Another possibility is that measures’ sensitivity to increased listening demands is influenced by how challenging the task is and by participants’ ability to cope with the demands of the task.

Although alpha power is interpreted as an index of listening effort, it can sometimes be difficult to interpret what a decrease in alpha power indicates. For example, Petersen and colleagues (2015) suggested that individual differences in cognitive capacity make it difficult to determine whether a drop in alpha power is an indication of decreased listening effort or an indication of exhausted cognitive resources, especially in cases where the accuracy of task performance is consistent across different listening conditions or groups of participants.

The sensitivity of alpha activity to changes at the level of the individual participant is another issue. Dimitrijevic et al. (2017) found that speech recognition in background noise evoked greater alpha power in adverse SNRs than in favourable ones. Increases in alpha power were statistically significant at the group level, but were identified in only 4 out of 14 participants. With the exception of Dimitrijevic et al. (2017), most studies have focused on average alpha power across participants. Alpha power may not be a reliable index of listening effort at the individual level.

*Event Related Potentials (ERP)*

Obleser and Kotz (2011) investigated the effect of context and degradation of the speech signal on the amplitude of N100 and N400 components using ERP. N100 and N400 are
negative potentials that should be normally observed at 100 and 400 ms, respectively, after presenting an auditory signal. The effect of context was assessed by presenting participants with “cloze sentences”, i.e. sentences consisting of a pronoun, verb, and object and are differentiated as high cloze (easy condition; object can be predicted from the context of the sentence, e.g. she sifts the flour) or low cloze (difficult condition; object cannot be predicted from the context of the sentence, e.g. she weighs the flour). Acoustic degradation of the sentence was achieved using a vocoding technique. Participants were instructed to rate the comprehensibility of the sentences by selecting one of the multiple choices ranging from one to four (where one indicates that the sentence is not at all comprehensible and four indicates that the sentence is very comprehensible). The authors found that degraded signals yielded a shorter latency and larger amplitude of the N100 potential. Changes in the amplitude and the latency of N100 were proportionate to the amount of acoustic degradation of the signal, with greater acoustic degradation yielding larger amplitude and longer latency N100. Therefore, the authors considered changes in N100 a marker of “neural effort” in contrast to a mere onset/offset response that is likely to be consistent regardless of how degraded the signal was. The authors also reported larger amplitude of N400 when presenting participants with low cloze sentences and also interpreted this as increased listening effort.

The synchronisation of changes in ERP activity with the occurrence of the auditory stimulus suggested that changes in ERPs provide information about an aspect of attention that is related to orienting, i.e. involuntary shifting attention towards the occurrence of a novel stimulus (Pichora-Fuller et al. 2016). Changes in ERP have been considered as indications of listening effort because changes in task difficulty alter the magnitude of ERP
activity. However, more research is required to identify how the dimension of increased listening demands assessed by ERP relates to other measures of listening effort.

**Pupillometry**

Increased pupil size occurs as a result of increased activity in the autonomic nervous system in cases of increased attention, alertness, or arousal (Kramer 1991). Traditionally, increased pupil size in demanding listening conditions (such as the presence of background noise) has been interpreted as indicating increased listening effort. For example, Zekveld et al. (2010) had participants repeat sentences presented in background noise. Increased pupil size was observed as the level of background noise increased and was interpreted as a sign of increased listening effort. Koelewijn et al. (2012) have also reported that presenting participants with speech material in the presence of background noise that had contextual information (e.g. single-talker masker) resulted in increased pupil size compared to background that did not contain contextual information (e.g. stationary noise). Increased pupil size has been often considered an indication of a negative experience as it has generally been observed with increased task difficulty, i.e. with greater likelihood of ineffective effort (i.e. effort that is not perceived as achieving successful task performance). Recent findings suggest that increased pupil size may not be always considered an indication of a negative experience that is likely to be associated with self-reported effort. Evidence suggests that increased pupil size can sometimes be considered an indication of effective effort (that is, associated with improved ability to engage in the listening task) and not necessarily associated with self-reported effort, as will be discussed in the examples below.
Kuchinsky et al. (2014) used pupillometry to measure changes in listening effort associated with speech training. The hypothesis was that training results in decreased listening effort and thus decreased pupil size. Speech training resulted in increased mean pupil size (compared to pupil size before the training), along with improved performance in the speech test. The authors suggested that the increase in pupil size might be an indication of increased task engagement associated with successful performance, i.e. effective allocation of cognitive resources.

Wendt et al. (2016) used pupillometry to measure listening effort associated with performing a speech comprehension task. Participants also performed tests of WM capacity. It was hypothesised that participants with larger WM capacity would experience less listening effort than participants with smaller WM capacity and would therefore exhibit smaller pupil sizes and report less listening effort. The researchers found that participants with larger WM capacity had larger pupil size and reported less listening effort compared to participants with less WM capacity. Wendt and colleagues suggested that increased pupil size might be an indication of increased ability to attend and engage in the task, due to better ability to cope with its demands. They also suggested that self-report measures of listening effort and pupillometry assess different aspects of listening effort and that processing effort does not necessary translate into perceived effort. The findings of Kuchinsky et al. (2014) and Wendt et al. (2016) suggest that increased pupil size with improved task performance can indicate better ability to attend to the task and cope with its demands, i.e. effective effort.
To test the hypothesis that performance can influence participants’ experience of listening effort, Ohlenforst et al. (2017b) measured listening effort in groups of adults with hearing impairment and adults with normal hearing when presented with speech at different SNRs. The aim was to identify how peak pupil size varied with changes in performance. Ohlenforst et al. (2017b) examined peak pupil size when participants with normal hearing and participants with hearing impairment listened to sentences in the presence of two types of background maskers (single talker and stationary background noise) at different SNRs. A significant interaction was identified between participant group and both SNR and type of masker. In stationary background noise, peak pupil size for participants with normal hearing was identified at a narrow range of unfavourable SNRs. In addition, changing the levels of background noise resulted in significant changes in peak pupil size; i.e. peak pupil size increased with higher levels of background noise up to the point where participants were unable to cope with the demands of the task and a drop in peak pupil size (and performance) was observed. The pattern of change in peak pupil size in participants with normal hearing is consistent with the Framework for Understanding Effortful Listening [FUEL; (Pichora-Fuller et al. 2016)], which suggests that inability to cope with the demands of the task results in decreased motivation and in a drop in listening effort (further discussion of the FUEL is provided in section 2.4).

In participants with hearing impairment, peak pupil size did not change significantly across SNRs. The authors suggested that hearing-impaired participants had already recruited significant cognitive resources for the relatively favourable SNRs. Consequently, there was minimal opportunity for further resources to be used in the challenging listening conditions. Similar pattern of change in peak size was identified in both groups
when listening to sentences in the presence of a single talker masker with a shift in the peak towards the more favourable SNRs in participants with hearing impairment. The authors suggested that different patterns of change in peak pupil size between the groups might suggest that participants with hearing impairment have different pattern of allocation of cognitive resources than individuals with normal hearing in everyday listening situations. The effect of performance on changes in pupil size might explain the findings of McMahon and colleagues, who found no difference in pupil size when presenting participants with 6- and 12-channel vocoded sentences in challenging listening conditions (≤ 0 SNR). The absence of a difference might be a result of inability to cope with the demands of the task, as evidenced by the significant drop in participants’ performance at these SNRs.

Kramer et al. (2016) identified increased peak pupil size in participants with normal hearing and participants with hearing impairment when performing a sentence repetition in noise task at 50% criterion performance (difficult condition) compared to performing the task in quiet (easy condition). No difference in peak pupil size was identified between the groups in the easy listening condition. However, participants with normal hearing had larger peak pupil size than participants with hearing impairment in the difficult listening condition. The authors suggested that participants with hearing impairment are likely to expend increased cognitive resources to extract speech from background noise compared to participants with normal hearing. Therefore, participants with hearing impairment would have less spare cognitive capacity available for task performance. The authors suggested that the difference in spare cognitive capacity available for task performance might have contributed to the increased pupil size observed in participants with normal
hearing. The findings of Kramer and colleagues provide another example of cases where increased effort can be considered effective, as it was associated as indication of better ability to attend to the task.

The findings of the studies discussed above suggest that changes in pupil size are influenced by task demands, listeners’ performance, available cognitive resources, and motivation to engage in the listening task. The influence of multiple factors on pupillometry might have contributed to the often-reported absence of correlation between self-report measures of listening effort and pupillometry, e.g. Zekveld et al. (2010) and Koelewijn et al. (2012). Maximum pupil size is likely to be observed when individuals perform tasks of moderate difficulty (Ohlenforst et al. 2017b). As the level of difficulty increases, a decrease in pupil size is likely to be observed. When the challenges posed by a task become so great that they cause the participant to ‘give up’, pupil size may drop so that it has the same size observed while performing the task at a mild level of difficulty. Therefore, interpreting the findings obtained using pupillometry without considering aspects of performance or perceived effort might lead to inaccurate conclusions. The influence that performance is likely to have on changes in pupil size highlights the importance of establishing a psychometric function for the performance of each participant when using pupillometry as a measure of listening effort, e.g. a decrease in pupil size that is associated with impaired performance is an unlikely indication of effective effort (where individuals are able to cope with demands of the task without exerting much effort). However, establishing a psychometric function for each patient in clinical settings might be unfeasible, given time constraints.
Measures of changes in the autonomic nervous system (skin conductance, skin temperature, electromyographic activity, heart rate)

The autonomic nervous system consists of a sympathetic and a parasympathetic branch. The sympathetic nervous system is responsible for the “fight or flight” response in the body, i.e. responses to what the body perceives as an emergency. The activity of the sympathetic nervous system usually increases in response to stressful events and results in bodily changes such as increased heart rate and accelerated breathing. The parasympathetic branch of the autonomic nervous system is responsible for recovery responses associated with termination of stressful events. The parasympathetic nervous system has opposing effects to those of the sympathetic nervous system, e.g. it results in decreased heart rate and decelerated breathing rate (McArdle et al. 2006). Increased mental load results in increased activity in the sympathetic branch of the ANS and decreased activity in the parasympathetic branch of the ANS (Staal 2004). Increased sympathetic activity can result in changes in skin conductance, skin temperature, electromyographic (EMG), and heart rate variability (Andreassi 2013).

Mackersie and Cones (2011) investigated the effect of increased listening demands on skin conductance, skin temperature, heart rate, and EMG activity in participants with normal hearing. Subjective ratings of listening effort were measured using the NASA Task Load Index (Hart and Staveland 1988). A dichotic listening task was used where task difficulty was manipulated by varying the number of the digits presented to one or two ears (low, medium, and high demand listening conditions). There was a significant change in mean EMG and skin conductance with increased task difficulty. Z-scores referenced to the low demand listening condition were calculated for each participant to determine the
presence of significant change in any of the measures at the medium and the high
demand listening conditions. Any z-score greater than 2 was considered a significant
change from the low demand condition. Sixty percent of the participants showed
significantly increased skin conductance with increased task demands. However, due the
variability in EMG scores across participants, there were no significant changes in EMG
activity at the individual level suggesting that the applicability of EMG as a clinical
measure is questionable. The authors identified a weak correlation between skin
conductance and self-reported listening effort.

Mackersie et al. (2015) investigated the difference in listening effort between adults with
normal hearing and adults with hearing impairment using skin conductance, heart rate
variability, and a self-report measure. Participants performed a speech-in-noise task at
different SNRs. Increased skin conductance was identified in participants with hearing
impairment compared to participants with normal hearing across the different SNRs.
However, changes in the SNR did not result in significant changes in skin conductance.
Changes in heart rate variability were observed with changes in the SNR. However, a
difference in heart rate variability between the groups was only observed at the
challenging listening conditions (-6 and -3 SNR). There was no significant difference in
self-reported listening effort between the groups. The authors suggested changes in the
levels of background noise might not have been large enough to induce changes in skin
conductance levels. In Mackersie and Cones (2011), skin conductance only increased
when task difficulty changed from very low to medium but plateaued thereafter.
Results suggested that the pattern of change in the activity of the ANS is not consistent across different levels of task difficulty. In Mackersie et al. (2015), a difference in heart rate variability between the groups was observed only in the challenging listening conditions, while a difference in skin conductance between the groups was observed at all SNRs. The influence of both task load and emotional reactivity on measures of changes in the ANS complicates their use as indicators of listening effort. Any observed change in these measures could be due to task difficulty, emotional stress, or both (Mackersie et al. 2015). Mackersie and Calderon-Moultrie (2016) argued that changes in the response of the ANS can occur as a result of changes in aspects such as motivation or stress in response to factors related to the test environment, and are not necessarily a mere indication of increased listening effort associated with task performance. For example, increased skin conductance might be observed as a result of stress associated with perceived failure in task performance. The authors also suggest that stress associated with listening to speech in background noise is more likely to occur in hearing-impaired individuals, who also commonly report increased disturbance by environmental sounds compared to individuals with normal hearing. The limited responsiveness of skin conductance to changes in noise levels reported by Mackersie and Cones (2011) and Mackersie et al. (2015) might suggest that increased skin conductance is an indication of stress that is merely associated with the presence of noise and is independent of task difficulty. More research is required to investigate the reliability of skin conductance as a measure of listening effort, and also how the aspect of increased listening demands assessed by skin conductance relates to other potential measures of listening effort.
2.3.2 Measures of listening-related fatigue

2.3.2.1 Self-report measures of listening-related fatigue

Hornsby and Kipp (2016) compared the self-reported vigour and fatigue of two groups: adults seeking help for their hearing difficulties and a control group of adults assumed to have good hearing. Information about hearing aid use was not provided. Self-reported vigour and fatigue of the participants in the control group were obtained from normative data of healthy participants. Therefore, the authors assumed that participants in the control group had good hearing. Participants with hearing impairment completed the following self-report scales: i) the Multi-dimensional Fatigue Symptom Inventory-Short Form [MFSA-SF, (Stein et al. 2004)]; ii) a 15 item subscale from the Profile Of Mood States [POMS, (McNair 1971)] that assess aspects of fatigue and vigour; and iii) the Hearing Handicap Inventory for the Elderly/Adults.

Participants with hearing impairment reported less vigour than controls. However, the difference in self-reported fatigue between the groups was not significant. The presence of a hearing problem in a number of participants in the control group might have contributed to the absence of a difference in self-reported fatigue as the authors did not have information about their hearing thresholds. There was no correlation between hearing sensitivity and each of self-reported fatigue and vigour, which the authors suggested might be due multiple factors such as individual differences in speech perception abilities and in aspects as motivation and how rewarding the listening experience is. However, the authors identified a significant correlation between hearing disability and self-reported fatigue and vigour.
Hornsby and colleagues’ findings suggested that perceived success in task performance (indicated by perceived hearing handicap) might provide indications about individuals’ experience of fatigue. Individuals who report struggling in different social and emotional aspects of their life are likely to experience increased levels of fatigue. The findings of Hornsby and Kipp are consistent with the MCM (Hockey 2013) discussed earlier. The authors suggested that self-reported hearing handicap might be an important aspect to consider in hearing rehabilitation. Individuals with hearing impairment are likely to obtain increased benefits from a hearing rehabilitation strategy that considers aspects of motivation and perceived communicative success. More research is required to confirm the association between hearing handicap and self-reported listening effort and fatigue.

2.3.2.2 Behavioural measures of listening-related fatigue

Fatigue can be behaviourally assessed using an “indirect approach” and a “direct approach” (Hornsby et al. 2016). The indirect approach involves measuring cognitive abilities before and after a prolonged period of time (such as before and after a school day). In the indirect approach, the amount of mental effort exerted can vary across individuals. In contrast, the direct approach decreases individual variability by having individuals perform a similar task (Hornsby et al. 2016). Vigilance tasks are common examples of a behavioural direct approach for the assessment of fatigue (Hornsby et al. 2016). Vigilance tasks require participants to maintain attention to respond to the occurrence of simple infrequent events. The development of fatigue during task performance can interfere with individual’s ability to continue performing with the same
level of accuracy. The deterioration in performance over the duration of the task is the basis for the behavioural assessment of fatigue using vigilance tasks (Hornsby et al. 2016).

Hornsby (2013) used a behavioural method (dual-task paradigm) to investigate the effect of hearing aid use on listening effort and fatigue. The primary task involved single word repetition. Two secondary tasks were used that involved word memorising and a visual response time task. It was hypothesised that fatigue would develop over the duration of task performance (60 minutes) and that this would result in deterioration in the performance on either the primary or any of the secondary tasks. Results indicated a systematic deterioration in participants’ performance for the visual response time task only in the unaided condition. The author suggested that hearing aids were beneficial in terms of reducing listening-related fatigue. Hornsby did not identify a difference in self-reported fatigue between the aided and the unaided conditions. There was no correlation between self-report and behavioural measures or between hearing sensitivity and any of the measures. The absence of a correlation between the measures suggests they assess independent aspects of fatigue. Alternatively, fatigue might have not developed as indicated by the self-report measure. The drop in participants’ performance overtime might be due to other factors such as boredom. However, this does not explain the absence of a drop in the performance of the primary task or in the or in the secondary task that involved word memorising.

One of the limitations of using decrement in task performance as a measure of fatigue is that decrement in performance over time does not necessarily indicate the development of fatigue. It can occur as a result of lack of engagement or boredom (Hockey 2013).
Likewise, the absence of a decrement in task performance over time does not always imply that fatigue did not develop. Learning effects can sometime minimise any effect fatigue is likely to have on task performance (Hornsby et al. 2016). Additionally, effortful processes to maintain task performance can prevent any effect fatigue can have on task performance. Increased levels of effort to maintain task performance can be accompanied with a “fatigue after-effect” in which fatigue have an effect on the post-task activities but not on the task itself (Hockey 2013). According to Hockey (2013), any measured behavioural or physiological change is of minimal value if fatigue was not reported by the individual.

2.3.2.3 Physiological measures of listening-related fatigue

Cortisol is a hormone that is produced by the adrenal gland (Kirschbaum et al. 1996). The development of fatigue can sometimes result in decreased production of cortisol (Nemeroff 1998). Hicks and Tharpe (2002) compared cortisol levels in children with normal hearing and with hearing impairment at the beginning of one school day and at the end of another. Cortisol levels were significantly higher in the morning than in the afternoon for both groups. However, there was no significant difference in cortisol levels between the two groups. The authors suggested that the absence of a difference in cortisol levels between the groups might be because children with hearing impairment used hearing aids which might have reduced fatigue. Other possible reasons for the absence of difference between the groups include: i) children with hearing impairment had mild to moderate hearing loss, and changes in cortisol levels might only be sensitive to severe levels of hearing impairment; ii) cortisol levels were only measured twice a day, too infrequently to allow comparison of cortisol levels to the normal circadian cycle; and
iii) changes in cortisol levels might not be sensitive to the experience of fatigue induced by the presence of hearing impairment. Similarly, Kramer et al. (2016) found no significant difference in cortisol levels between adults with normal hearing and adults with hearing impairment when measured before, after, and in between performing two sentence recognition tasks (once in quiet and once in the presence of background noise). However the authors identified increased Chromogranin A levels (a protein produced as a result of the activation of the neuroendocrine system and used as a marker of stress) at pre-test compared to the other three conditions, which was interpreted as increased stress levels prior to task performance.

More recently, Bess et al. (2016) identified increased awakening cortisol levels in children with hearing impairment compared to children with normal hearing. 30 minutes after waking, decreased cortisol levels were identified in children with hearing impairment compared to children with normal hearing. Bess and colleagues suggested that increased awakening cortisol levels in children with hearing impairment might be an indication of increased stress and vigour in preparation for the start of the day. Decreased cortisol levels 30 minutes post-awakening was considered an indication of increased risk of fatigue in participants with hearing impairment. The authors argue that further research is required to identify how changes in cortisol levels relate to aspects of stress, fatigue, and energy control.

There are arguments that fatigue does not necessarily result in reduced cortisol levels (Cleare 2004; Kumari et al. 2009). However, conditions that cause chronic fatigue might lead to altered cortisol levels as a result of long term consequences of the condition (such
as lack of sleep); i.e. changes in cortisol levels might not be the result of fatigue *per se* (Cleare 2004; Kumari et al. 2009). Further work is required to identify the effects of hearing loss on cortisol levels.

Key et al. (2017) investigated fatigue resulting from performance of demanding speech-processing tasks (speech-recognition task, dual-task paradigm, and a speech-vigilance task) using self-report, behavioural, and physiological measures in children with normal hearing (aged 6 to 12.9 years old). Measures were obtained before and after participants performed the speech tasks. The self-report measure was a 5-point likert scale for rating current fatigue state. The behavioural measure was a sustained visual attention task, i.e. a vigilance task. ERP was used as a physiological measure of fatigue. The hypothesis was that fatigue resulting from performing the speech tasks would result in increased self-reported fatigue, deterioration in the performance of the behavioural (vigilance) task, and reduced ERP activity. As hypothesised, significant differences were identified in the measures before and after performing the speech tasks. Although the duration of task performance was 15 minutes only, the authors argued that their findings suggest the development of fatigue.

The authors investigated the relationship between the measures before the participants performed the speech tasks and also after speech tasks performance. The authors also computed before-after difference measures for each of the above measures, and tested for correlations between them. There was a significant correlation between increased ERP activity and increased response time in the vigilance task after participants performed the speech tasks. In addition, a greater reduction in the amplitude of P300 was
identified in participants who performed poorer in the behavioural task after controlling for the effect of age. The authors suggested that fatigue resulting from effortful listening can negatively impact cognitive functioning abilities. The findings need to be interpreted with caution, as the relationship was modest and there was considerable variability in the data suggesting that reductions in the amplitude of P300 were not associated with poorer performance in the behavioural task across all participants. The lack of significant correlation of each of the behavioural and the physiological measures with the self-report measure is unsurprising, given numerous research findings which suggest that self-report measures assess an independent aspect of listening effort or fatigue.

2.4 The underlying dimensions assessed by self-report, behavioural, and physiological measures of listening effort and fatigue

Inconsistencies in the findings across studies of listening effort raise questions about the underlying dimensions assessed by the different measures. The lack of correlation between the measures also raises questions about their reliability, as unreliable measures are unlikely to correlate with each other. The inconsistencies reported in the literature include: i) measures of “listening effort” that do not correlate with each other, ii) inconsistent results across studies that used the same measure, and iii) inconsistent patterns of results across different groups of participants.

Since there is no gold-standard measure of listening effort/fatigue, studies have often employed more than one measure, presumably in efforts to verify the sensitivity of the measures to the underlying phenomena of listening effort and fatigue. For example, self-
reported listening effort associated with task performance has been assessed in large number of studies alongside other behavioural and physiological measures, probably because it seemed intuitive to assume that behavioural or physiological indications of increased listening effort should be accompanied by a self-reported state of effort. As mentioned earlier, recent evidence has often suggested the possibility of the independency of the aspects assessed by the different measures, e.g. the exertion of increased cognitive load has not always translated into a self-reported state of listening effort. A number of models and frameworks have been developed to explain how listening effort and fatigue are influenced by multiple factors that are not limited to task difficulty such as the FUEL (Pichora-Fuller et al. 2016) and the MCM (Hockey 2013). Within these models and frameworks, it is assumed that the factors that influence listening/fatigue effort might not affect the different measures in a similar way, leading to the inconsistencies and the lack of correlations identified. A number of these models will be discussed below.

Pichora-Fuller et al. (2016) have elaborated on Kahneman’s capacity model (1973) in the FUEL. Kahneman’s capacity model suggests that increased task demands leads to increased effort. However, based on the FUEL, increased demands do not necessarily result in increased effort when there is low motivation, particularly if increased effort does not result in perceived successful performance. Individuals might not exert effort if motivation is low, even if the limit of cognitive capacity has not been exceeded.

An individual’s decision to exert effort can also be influenced by their appraisal of the task’s demands and the appraisal of their capacity to perform the task (Pichora-Fuller...
The arguments provided by the FUEL are supported by the findings of studies that report decreased listening effort when motivation to engage in the task is reduced as a result of impaired performance, e.g. Ohlenforst et al. (2017b). Consistent with the MCM (Hockey 2013), the FUEL suggests that fatigue is more likely to be reported in cases where increased effort is not perceived as resulting in successful task performance. Pichora-Fuller et al. (2016) argue that some measures used in the assessment of listening effort might be more sensitive than others to factors related to the demands of the task. For example, processing speech in background noise can affect measures of listening effort differently, e.g. McMahon et al. (2016). Other measures might be more sensitive to individual factors such as motivation, pleasure, or fatigue. In addition, some measures might be influenced by an interaction of both of the task demands and the individual factors, such as how motivated an individual is to engage in a conversation in a noisy environment. The inconsistent effects that the aforementioned factors have on the different measures could be a reason for the lack of correlation between them.

Consistent with the FUEL, Lemke and Besser (2016) have proposed the terms “perceived effort”, i.e. subjective estimation of how taxing a listening task is or was, and “processing effort”, i.e. the amount of processing resources allocated to the task. The authors suggested that processing effort associated with a task does not necessarily translate into perceived effort if the listener is motivated. The independency of the concepts of perceived effort and processing effort proposed by Lemke and Besser might explain the lack of correlation between self-report and behavioural/physiological measures of listening effort. Lemke and Besser have also suggested that factors that are not limited to task difficulty interact and affect perceived and processing effort in different ways.
Examples of these factors include: the physiological state of the participant (e.g. tiredness, sleepiness, and arousal), motivation, feelings of stress, and emotional state.

Task difficulty has often been considered the main influencer of any changes identified in the self-report, behavioural, or physiological measures of listening effort/fatigue. However, the factors suggested by Lemke and Besser (2016) also need to be taken into account when interpreting the results obtained using the different measures. For example, it is important identify how the multiple factors proposed by Lemke and Besser might have influenced the findings obtained using a certain measure, rather than assuming that changes in task demands are the main influencer of any changes observed in listening effort measures. As previously discussed, increased pupil size might not be always associated with self-reported effort if perceived success in task performance motivates further exertion of effort. It is also important to consider that measures have different sensitivity to emotional and psychological factors, which might contribute to the lack of correlation between them (Hogervorst et al. 2014).

Strauss and Francis (2017) have recently proposed a taxonomic model in an attempt to explain the multidimensional nature of listening effort. According to Strauss and Francis, effortful listening has both external and internal dimensions. An example of an external dimension is having to communicate with friends in a noisy environment. An example of an internal dimension is trying to understand the contents of a lecture on an unfamiliar topic. Strauss and Francis suggested that different measures of listening effort might have different sensitivity to the external and internal dimensions of listening effort and that this might have contributed to the absence of correlation and the inconsistent findings reported in the literature.
As discussed in the FUEL, Strauss and Francis suggest that the amount of effort imposed by the task, “demanded effort”, does not necessarily equal the amount of “exerted effort”. The authors propose the term “resource limiting”, which they define as the difference between the effort demanded by the task and the effort exerted by the participant. When a group of participants perform the same listening task, individual differences in aspects as motivation, arousal, and attentional control might result in differences in amount of effort exerted by the participants, despite the fact that demanded effort is similar for all participants. The difference in resource limiting across participants might be the reason for the individual differences in listening effort measured using behavioural or physiological methods.

Peelle (2017) has recently suggested that the cognitive processes required for understanding degraded speech are “domain general”. Evidence from neuroimaging studies suggests that multiple cognitive systems are activated when perceiving degraded speech. The cognitive processes involved in the comprehension of degraded speech (including verbal working memory and attention-based performance monitoring) can differ depending on the acoustic, cognitive, and linguistic demands of the task. Peelle’s arguments support the hypothesis that listening effort might be multidimensional. According to Peelle, individual differences in cognitive abilities can also result in differences in the cognitive processes engaged in processing degraded speech, which might also explain the lack of consistency in listening effort measures.

Evidence suggesting the multidimensionality of listening effort and fatigue has recently increased. The inconsistencies and the lack of correlation between potential measures
might thus be due truly to the multidimensionality of the concepts. However, it can also
due to the measures being unreliable, as unreliable measures are unlikely to correlate
with each other. To the candidate’s knowledge, the reliability of listening effort and
fatigue measures has not been previously investigated. An in-depth understanding of how
the different self-report, behavioural, and physiological measures of listening
effort/fatigue relate to each other and an assessment of their reliability are required. It is
also important to gain understanding of the underlying dimensions of listening effort/
fatigue that are assessed by the different measures. Investigating the pattern of
correlation between the measures requires simultaneous use of these measures during a
single listening task and the use of data reduction method in the analysis (such as Factor
Analysis) to determine whether the measures tap into similar or independent underlying
dimensions.

2.5 Summary and gap in knowledge

In this review, an overview of the concepts of listening effort and fatigue was first
provided, along with a discussion of models and theories that explain these concepts. The
first gap in knowledge identified was the absence of an attempt to quantify self-reported
listening effort and fatigue in adults with hearing impairment. Increased self-reported
listening effort and fatigue in individuals with hearing impairment would justify
considering these phenomena in the assessment of hearing disability. There have also
been few attempts to investigate how self-reported listening effort and fatigue correlate
with hearing levels and other aspects of perceived hearing difficulties or hearing
handicap. Models and theories suggest the presence of an association between perceived
difficulties and both effort and fatigue (Hockey 2013). Identifying a correlation between perceived difficulties or handicap and both listening effort and fatigue would have important implications for hearing rehabilitation. This review has also discussed a number of potential measures of listening effort and fatigue. Although various measures purport to assess the same construct, measures have often failed to agree with each other and have sometimes contradicted one another. The lack of agreement between measures raises questions about their reliability, sensitivity, and the underlying dimensions assessed by each of them.

The first aim of this thesis was to systematically quantify self-reported listening effort and fatigue in individuals with hearing impairment. It was hypothesised that individuals with hearing impairment would report significantly increased levels of listening effort and fatigue compared to a matching control group with good hearing. The finding that individuals with hearing impairment experience increased levels of listening effort and fatigue would justify the importance of identifying a clinically relevant assessment method. A second aim of the thesis was to identify correlates of self-reported listening effort and fatigue, by investigating their relationship with aspects of hearing sensitivity, hearing handicap, and speech recognition abilities. Based on the findings of Hornsby and Kipp (2016), and on the models and frameworks that discuss the relationship between aspect of effort, fatigue, and perceived difficulty (Hockey 2013; Pichora-Fuller et al. 2016), it was hypothesised that hearing handicap would have the strongest correlation with perceived listening effort and fatigue. The final aim for this research was to investigate the reliability of proposed measures of listening effort, how they correlate with each other, and whether they tap into the same underlying concept. The overall aim of the
thesis was to improve our understanding of the underlying dimensions assessed by the different measures, because this could have widespread implications for clinical and research purposes, such as identifying whether a group of measures or a single measure is required for the assessment of listening effort or fatigue in clinical or research settings.
CHAPTER THREE

STUDY ONE: SELF-REPORTED LISTENING-RELATED EFFORT AND FATIGUE IN HEARING-IMPAIRED ADULTS

This manuscript has been accepted for publication in *Ear and Hearing*:


The format of *Ear and Hearing* manuscripts is used in the chapter.

Page number of thesis: 76
Self-Reported Listening-Related Effort and Fatigue in Hearing-Impaired Adults

Sara Alhanbali,1 Piers Dawes,1 Simon Lloyd,2 and Kevin J. Munro1,3

Objective: Hearing loss may increase listening-related effort and fatigue due to the increased mental exertion required to attend to, and understand, an auditory message. Because there have been few attempts to quantify self-reported effort and fatigue in listeners with hearing loss, that was the aim of the present study.

Design: Participants included three groups of hearing-impaired adults: (1) hearing aid users (HA, n = 50; 31 male, 19 female; age range = 55 to 85 years); (2) cochlear implant users (CI, n = 50; 26 male, 24 female; age range = 55 to 80 years); and (3) single sided deafness (SSD, n = 50; 30 male, 20 female; age range = 58 to 80 years). There was also a control group of adults who passed a hearing screen at 30 dB HL at the frequencies: 500, 1000, 2000, and 4000 Hz in both ears (n = 50; 22 male, 28 female; age range = 55 to 78 years). The fatigue assessment scale (FAS) was used to quantify fatigue. The FAS is a generic standardized self-report scale consisting of 10 items that are scored using a five-point Likert scale. An effort assessment scale (EAS), developed for the present study, consisted of six questions with responses provided on a visual analog scale that ranges from 0 to 10.

Results: All hearing-impaired groups reported significantly increased effort and fatigue compared to the control group. The median fatigue score for the control group was 14 and around 22 for the three hearing-impaired groups. The median effort score for the control group was 20 and around 70 for the three hearing-impaired groups. There was no significant difference in mean effort or fatigue between the three groups of hearing-impaired adults. There was a weak positive correlation between fatigue and effort scores (r = 0.40, p < 0.05). The proportion of participants with extreme fatigue (scores above the 95th percentile of the control group) was 22, 10, and 22%, for the HA, CI, and SSD groups, respectively. The proportion of those with extreme effort was 46, 54, and 52%, for the HA, CI, and SSD groups, respectively. Results of factor analysis using the individual questions from both questionnaires indicated that the questions loaded into two factors: a “fatigue” factor for all of the FAS questions and an “effort” factor for all of the EAS questions.

Conclusion: Hearing-impaired individuals report high levels of listening effort and fatigue in everyday life. The similarity in listening-related effort and fatigue between the different hearing-impaired groups suggests that these aspects of listening experience are not predicted by the severity of hearing impairment. Factor analysis suggests that the FAS and the EAS assess two distinct dimensions. The low correlation between FAS and EAS means that fatigue cannot be reliably predicted from self-reported effort in individual listeners.

Key words: Fatigue, Listening effort, Self-report scales.

(Ear & Hearing 2017;38;e39–e48)

INTRODUCTION

A recent discussion paper provided a general definition of listening effort as “the mental exertion required to attend to, and understand, an auditory message” (McGarrigle et al. 2014, p. 434). This general definition does not seek to differentiate processing effort from perceived effort; however, the present study is concerned with the latter, the self-reported effort associated with listening. The same discussion paper also provided a general definition for listening-related fatigue as “the extreme tiredness resulting from effortful listening” (McGarrigle et al. 2014, p. 434). In previous research on listening effort and fatigue, a variety of self-report, performance-based, and physiological measures have been used (see reviews by Bess & Hornsby 2014; McGarrigle 2014). Implicit in this body of research is the idea that listening effort and listening-related fatigue have a cognitive basis that has physiological correlates; however, the relationship between the physiological indices of listening effort, and the self-reported listening effort (processing and perceived effort, respectively) is complex and is not well understood (e.g., see Wendt et al. 2016).

Hearing-impaired listeners may expend increased listening effort in difficult listening situations compared to normal-hearing listeners. Increased listening demands are imposed on hearing-impaired listeners in order for them to compensate for their hearing loss. For instance, a hearing-impaired listener might not be able to hear every single word in a sentence. Consequently, more mental effort may be required to identify the relationship between the different items in the sentence, guess misheard words, and the gist of the sentence. Increased listening effort might benefit hearing-impaired individuals in terms of understanding speech in challenging listening situations (Downs 1982; Hick & Tharpe 2002; Zekveld et al. 2011; Hornsby 2013). However, high levels of listening effort on a daily basis may result in mental fatigue, which may be associated with a reduced ability to concentrate or to perform some cognitive tasks (Hornsby 2013; Bess & Hornsby 2014). The tiredness resulting from increased listening effort could cause a hearing-impaired individual to “give up” on exerting effort to understand speech and this may lead to communicative disengagement (Hétu et al. 1993).

Hearing-related disability, that is, the listening difficulties associated with the presence of hearing impairment that induces limitations on the individual’s ability to function in everyday life, is currently measured using tests of speech perception and self-reported hearing disability. Performance on speech perception tests does not index listening-related effort or fatigue, so may miss an important dimension of hearing disability. Most of the self-report measures of hearing used in audiology do not include items about listening-related effort and fatigue. One exception is the speech, spatial, and qualities (SSQ) hearing scale (Gatehouse & Noble 2004), which contains three items about listening effort. However, the SSQ hearing scale is not commonly used in clinical settings. Quantifying listening effort and fatigue may provide a more detailed assessment of hearing-related disability and may act as a useful outcome measure when comparing...
intervention strategies. The aim of the present study was to quantify self-reported levels of listening effort and fatigue in hearing-impaired adults.

Below, we first discuss the concept of fatigue and its assessment in chronic health conditions other than hearing loss. We then discuss the relevance of fatigue to hearing-impaired listeners, and the importance of the assessment of fatigue as a dimension of hearing disability. We then discuss the concept of listening effort, its relevance in the field of audiology, self-report scales of listening effort, and their application.

Measuring Fatigue in Chronic Health Conditions

Fatigue is a commonly reported experience in a number of chronic health conditions (Dittner et al. 2004). In each health condition, the fatigue may be either physical, mental, or both (Lou et al. 2001); physical fatigue is defined as a reduced ability or desire to perform a physical task (Bess & Hornsby 2014) whereas mental fatigue is defined as a feeling of tiredness, exhaustion, or lack of energy due to cognitive or emotional demands (Bess & Hornsby 2014). Fatigue can reduce quality of life in terms of decreased productivity and increased work-related injuries (Ricci et al. 2007). Fatigue is also associated with depression and lack of desire to engage in daily life activities and social interactions (Ferrando et al. 1998). These negative consequences of chronic fatigue have raised awareness of the importance of its assessment in a number of chronic health conditions (Dittner et al. 2004).

Fatigue scales developed for specific patient populations consist of questions that target the symptoms that occur with the health condition (Dittner et al. 2004). Examples of these scales are the “cancer fatigue scale” (Okuyama et al. 2000) and one for patients with multiple sclerosis (MS) “MS-specific fatigue severity scale (FSS)” (Krupp et al. 1995). For instance, one of the items in the MS-specific FSS is “heat brings on my fatigue.” This is specific to MS because feeling weak as a result of heat is a commonly reported symptom in MS patients (Nelson & McDowell 1959).

Fatigue scales have been used to (a) assess the presence and the severity of fatigue, (b) estimate the number of “cases” who have severe fatigue, and (c) quantify the benefit of intervention in terms of reduced fatigue. For instance, Stone et al. (2000) aimed to identify the number of cancer patients experiencing severe fatigue. The authors defined cases of severe fatigue as those who obtained scores in excess of 95th percentile of the scores obtained by the control group on the FSS (Krupp et al. 1995). Results indicated that the percentage of patients complaining of severe fatigue was 15% for patients with recently diagnosed breast cancer, 16% for patients with recently diagnosed prostate cancer, 50% for patients with inoperable nonsmall cell lung cancer, and 70% for patients receiving specialist inpatient palliative care.

In addition to the disease-specific scales mentioned above, a range of fatigue scales have been developed for the general population including the Multidimensional Fatigue Inventory (MFI) developed by Smets et al. (1995). MFI consists of five subscales assessing multiple aspects of fatigue including: general fatigue, physical fatigue, mental fatigue, reduced motivation, and reduced activity. Some of the general fatigue scales mainly assess physical fatigue such as the FSS developed by Krupp et al. (1989), whereas others assess physical and mental fatigue such as the fatigue assessment scale (FAS) developed by Michielsen et al. (2004).

Hearing Loss and Fatigue

There are numerous anecdotal reports that the increased listening demands of hearing-impaired listeners cause them to feel tired and lacking in energy at the end of the day. The experience of hearing loss-induced fatigue on a daily basis can have negative long-term consequences affecting the quality of life of the hearing-impaired individual. For instance, Kramer et al. (2006) found that hearing-impaired workers tend to take more sick leave compared to their normal-hearing colleagues. Hearing-impaired workers reported that “fatigue” and “mental distress” are common causes for their frequent sick leave. Nachtegaal et al. (2009) have also found that hearing-impaired workers experience increased levels of fatigue at work and thus need more time to recover.

Unlike other chronic health conditions, self-reported fatigue within the hearing-impaired adult population has received limited attention. There are currently no self-report scales that have been specifically developed to assess self-reported listening-related fatigue. Previous studies have attempted to use general scales to index fatigue in groups of people with hearing loss. Hornsby et al. (2014) used the Pediatric Quality of Life Inventory (PedsQL) Fatigue Scale to investigate the difference in self-reported fatigue between a group of school-age children with hearing loss and a normal-hearing control group. PedsQL is a general self-report fatigue scale for children aged 5 to 18 years. PedsQL consists of three subscales assessing general fatigue (e.g., “I feel tired”), sleep/rest fatigue (e.g., “I rest a lot”), and cognitive fatigue (e.g., “It is hard for me to think clearly”). Hearing-impaired children reported significantly increased levels of fatigue in all three subscales compared to the control group.

Hornsby and Kipp (2016) measured fatigue in hearing-impaired adults >55 years of age who were seeking help for their hearing difficulties (the authors assumed that a small proportion of these adults might be existing HA users). A fatigue subscale from the Profile of Mood States (POMS; McNair et al. 1971) was used in the study. POMS consists of 65 single words that describe general feelings such as “anxious” and “energetic.” Six different mood states can be derived from the POMS including: “tension,” “depression,” “anger,” “confusion,” “fatigue,” and “vigor.” Data for an age-matched control group was obtained from a standardized sample of the general population (Nyenhuis et al. 1999). Hearing-impaired adults reported significantly less vigor compared to the control group. However, there was no significant difference in self-reported fatigue between hearing-impaired adults and the control group. Hornsby and Kipp (2016) did not have information about the hearing of the age-matched control group. Some of the participants in the control group might have had a hearing impairment, which could have reduced the difference in mean scores between the two groups.

In Hornsby and Kipp’s (2016) study, 15% of hearing-impaired adults reported severe fatigue, defined as scores that were more than 1.5 SD above the mean of the normative data. There was a relationship between the severity of fatigue and scores on the Hearing Handicap Inventory for Elderly (Ventry & Weinstein 1982), which is a measure of self-reported hearing
difficulties. However, no relationship was identified between the severity of fatigue and hearing level. Hornsby and Kipp suggest that the lack of correlation might indicate that perceived hearing difficulties are a stronger indicator of fatigue than hearing level.

To assess self-reported fatigue in listeners with hearing loss, it is necessary to use a validated general fatigue scale because no fatigue scales have been specifically developed to assess listening-related fatigue. Further, given that hearing loss is associated with increased cognitive rather than physical demands, it is reasonable to use a generic scale that includes items related to mental fatigue. One widely used general fatigue scale, with established reliability and validity, is the FAS (Michielsen et al. 2004). The FAS consists of items that assess both physical and mental fatigue. Smith et al. (2008) used FAS to investigate fatigue in elderly adult patients with stroke, chronic heart failure, and a control group. Stroke and chronic heart failure patients reported significantly higher levels of fatigue compared to the control group. The authors calculated the prevalence of “greater fatigue” by identifying the percentage of participants who obtained scores above the highest quintile of the scores obtained by the healthy control group. The prevalence of “greater fatigue” within stroke and chronic heart failure patients was 61.3 and 67.3%, respectively.

Measuring Listening Effort

Unlike fatigue, interest in the concept of effort is relatively limited in chronic health conditions. This may be because fatigue can be a chronic state, whereas effort is transient.

Hearing-impaired listeners commonly complain of the need for increased levels of effort to understand speech in background noise (McGarrigle et al. 2014). Measures of listening effort used in previous studies include self-report (e.g., Gatehouse & Noble 2004), performance-based/behavioral measures (e.g., reaction times; Gatehouse & Gordon 1990), and physiological indices (e.g., skin conductance and muscle activity; Mackersie & Cones 2011). However, self-report measures of effort do not generally correlate with behavioral or physiological measures of effort (e.g., Zekveld et al. 2010; Desjardins & Doherty 2013). This may be due, at least in part, to conflating processing effort and perceived effort: the relationship between these is complex and not well understood.

Kuchinsky et al. (2014) found that speech perception training resulted in improved word identification but this was accompanied by larger pupil sizes. Pupil dilation reflects increased vigilance and attention (Laeng et al. 2012) but it is counterintuitive to assume that the training resulted in higher levels of perceived effort. Wendt et al. (2016) found that varying the syntactic complexity of sentences and the level of background noise do not have the same effect on pupil dilation (processing effort) and self-report measures (perceived effort). They also found that participants with high working memory capacity showed increased pupil dilation in the higher-level noise condition but these same participants provided lower subjective ratings of listening effort in this condition. Therefore, it should not be assumed that pupil dilation and subjective ratings assess the same aspect of listening effort. In the clinical setting, it is the perceived effort that is of interest to the patient and the health care professional.

Some self-report measures have focused on the effort required to perform a specific listening task in an experimental or clinical setting. Mackersie and Cones (2011) used the National Aeronautics and Space Administration Task Load Index (NASA TLX) (Hart & Staveland 1988) to measure self-reported listening effort in relation to listening tasks of different levels of difficulty. Self-assembled questionnaires have been used to assess the listening effort experienced while performing a specific laboratory-based listening task. For example, Zekveld (2011) asked participants to rate the perceived listening effort on a scale ranging from 0 (no effort) to 10 (very high effort) in a sentence recognition in noise task at 50, 71, and 84% intelligibility levels.

Other self-report measures focus on listening effort that patients experience on a daily basis. Dawes et al. (2014) reported a “listening effort” subscale derived from three effort-related items in the SSQ hearing scale (Gatehouse & Noble 2004). Dawes et al. used this “listening effort” subscale to assess the change in listening effort following auditory acclimatization in new HA users compared with a control group of experienced HA users. Participants were required to rate the level of effort on a visual analog scale ranging from −5 to +5 with 0 indicating no change in effort, −5 much less listening effort and, +5 much more effort. After 3 months of hearing aid use, new HA users showed a significant reduction in listening effort with their new hearing aids compared with the control group.

To our knowledge, self-report measures have not been previously used to estimate the listening effort in the daily life of adults with hearing loss versus controls with good hearing.

Aims

This aim of this study was to extend previous knowledge by investigating both self-reported listening effort and self-reported fatigue in adults with different types of hearing impairment and to compare them with an age-matched control group with good hearing. Individuals with hearing loss included hearing aid users (HA), cochlear implant users (CI), and adults with single-sided deafness (SSD). It was hypothesized that individuals with hearing loss would report increased levels of listening effort and fatigue compared to the control group. A second aim was to investigate the relationship between the self-reported levels of effort and fatigue. If listening effort results in increased fatigue, it was hypothesized that there would be a positive correlation between the self-report levels of listening effort and fatigue. The last aim was to use factor analysis to investigate if the questions of the FAS and the effort assessment scale (EAS) load into two distinct dimensions, consistent with “effort” and “fatigue” being different constructs. It was hypothesized that the questions of the FAS would load into a “fatigue” factor, and the questions of the EAS would load into an “effort” factor.

MATERIALS AND METHODS

Participants

Four groups of 50 English speaking adults were recruited. Demographic data for each group of participants are provided in Table 1. Kruskal–Wallis test showed no significant difference in age between the groups \((H[3] = 6.066, p > 0.05)\). A minimum sample size of 40 participants per group was estimated to provide 80% statistical power to detect a clinically significant difference with a medium-sized effect \((r = 0.3; Field 2009)\) between the groups \((\alpha = 0.05)\), based on a between-groups analysis of variance. The study was powered to detect at least a
medium-sized effect because small difference between hearing-impaired and control groups would not be of clinical relevance. According to Field, the cutoff for a small effect size = 0.1 and the cutoff for a large effect size = 0.9. Power calculation was performed using G* power calculator version 3.1.

The HA group included adults with bilateral mild-to-severe sensorineural hearing loss. All of the participants in this group were users of one or two hearing aids for at least six months (16% were users of one hearing aid and 84% were users of two hearing aids). The CI group included adults who were users of one cochlear implant for at least six months. The SSD group included adults with profound unilateral hearing loss caused by the surgical removal of an acoustic neuroma at least one year earlier. All of the participants in the SSD group had hearing thresholds <35 dB HL average at 500, 1000, 2000, and 4000 Hz in the nonaffected side. The control group included adults who passed a pure-tone screen at a level of 30 dB HL at 500, 1000, 2000, and 4000 Hz in both ears.

Participants in the four groups were matched for age because of the evidence that age has an influence on listening effort (Degeest et al. 2015). Participants were also matched for sex. We did not match participants for socioeconomic status or educational level as we were not aware of any evidence to suggest that these factors would influence participants’ rating of listening effort and fatigue.

The study was reviewed and approved by the National Research Ethics Services of South Central—Hampshire A, REC reference: 15/SC/0113.

Self-Report Scales

Fatigue • The FAS is a validated scale consisting of 10 short items (Michielsen et al. 2004; see Table 2). Responses are provided on a five-point Likert scale with zero points for “never” and four points for “always.” The instructions were “The following 10 statements refer to how you usually feel on a daily basis. For each statement, choose one out of the five answers. Please give an answer to each statement, even if you do not have any complaints at the moment.” The overall score of FAS is calculated by summing the responses obtained to each individual question. The total score of FAS ranges from 0 to 40, with higher scores indicating more fatigue. To investigate the correlation between FAS and EAS, FAS scores were converted into percentages.

Effort • We are not aware of any validated scale to measure self-reported listening effort in the daily life of people with hearing loss. Consequently, we chose to use a self-assembled scale. We refer to this scale as “EAS” (Table 3). Three of the EAS questions were obtained from the SSQ hearing scale (Gatehouse & Noble 2004), which is a validated scale assessing different aspects of hearing disability. The other three questions were from an unpublished PhD paper (Alkhamra, Reference Note 1).

In the EAS, Responses are provided on a visual analog scale from 0 to 10 with 0 indicating “no effort” and 10 “lots of effort.” Participants are required to put a mark at the point that represents the level of effort they experience. The total score of EAS was calculated by adding the score of each of the six questions to give a score between 0 and 60, with higher scores indicating more effort. As in the case of FAS, all scores were converted into percentages.

Procedure

Hearing-Impaired Groups • In each recruitment site, audiologists identified potential participants who met the inclusion criteria by reviewing the hospital records. For each hearing-impaired group, the questionnaires were posted initially to 80 potential participants along with an invitation letter, participant information sheet, consent form, and a stamped addressed envelope. Additional participants were also approached through the same recruitment sites to achieve a total of 50 participants in each group. In the invitation letter, participants were asked to complete the questionnaires, sign the consent form, and post them back to the researcher if they would like to take part in the study. Demographic and audiometric data (based on the most recent audiogram obtained within 3 months of participation in the study) were obtained from the patient records.

Control Group • Participants in the control group were approached directly through social groups. After written informed consent, the hearing level of the participants in the control group was checked to determine their candidacy to take

### Table 1. Summary data for each group of participants

<table>
<thead>
<tr>
<th>Group</th>
<th>Median Age, Years, with Range in Parenthesis</th>
<th>Male with Percentage in Parenthesis</th>
<th>Range of Hearing Thresholds in the Better Ear</th>
<th>Range of Hearing Thresholds in the Poorer Ear</th>
</tr>
</thead>
<tbody>
<tr>
<td>HA (n = 50)</td>
<td>72 (55–85)</td>
<td>31 (62%)</td>
<td>40–100 dB HL</td>
<td>40–100 dB HL</td>
</tr>
<tr>
<td>CI (n = 50)</td>
<td>71 (55–80)</td>
<td>26 (52%)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>SSD (n = 50)</td>
<td>68 (58–80)</td>
<td>30 (60%)</td>
<td>25–30 dB HL</td>
<td>NA</td>
</tr>
<tr>
<td>Control group (n = 50)</td>
<td>71 (55–78)</td>
<td>22 (54%)</td>
<td>NA (participants passed hearing screening at 30 dB HL at 500, 1000, 2000, and 4000 Hz)</td>
<td>NA (participants passed hearing screening at 30 dB HL at 500, 1000, 2000, and 4000 Hz)</td>
</tr>
</tbody>
</table>

### Table 2. FAS questions

1. I am bothered by fatigue
2. I get tired very quickly
3. I do not do much during the day
4. I have enough energy for everyday life
5. Physically, I feel exhausted
6. I have problems starting things
7. I have problems thinking clearly
8. I have no desire to do anything
9. Mentally, I feel exhausted
10. When I am doing something, I can concentrate quite well

FAS, fatigue assessment scale.
part in the study. The researcher visited social groups and performed hearing screening for potential participants before having them complete the questionnaires. Hearing screening was carried out using a Kamplex KLD 21 diagnostic audiometer. A pure tone was presented at a fixed level of 30 dB HL at the following frequencies: 500, 1000, 2000, and 4000 Hz. A “pass” was defined as being able to hear the tones at all frequencies in both ears. Only adults who passed hearing screening were included in the control group.

Analysis

The data were examined using Shapiro–Wilk and Kolmogorov–Smirnov tests and the findings indicated that it was appropriate to use nonparametric statistics for data analysis. Data were summarized using medians and percentiles. A comparison of scores between groups was carried out using a Kruskal–Wallis test. Post hoc analysis using Mann–Whitney U pair-wise test was carried out to identify any significant difference between any two groups. Bonferroni correction was applied (0.05 divided by 6) so all effects are reported at a 0.008 level of significance. The effect size was calculated by dividing the z score by the square root of the number of the participants included in the comparison (100 participants for each pair-wise comparison). The relationship between listening effort and fatigue was analyzed using Spearman’s correlation coefficient. We also calculated the proportion of participants who experience “extreme” listening effort and fatigue. The reference range was defined as scores above the 95th percentile of the control group. Chi-square test was used to identify any significant difference in the proportions of extreme effort and fatigue between the hearing-impaired groups. The Kaiser–Meyer–Olkin (KMO) measure was conducted to verify the adequacy of the sample size for factor analysis. Factor analysis was conducted on 16 items (10 questions of the FAS and 6 questions of the EAS). The factors were identified based on eigenvalues greater than one. Oblique rotation was applied to identify how the question of EAS and FAS load into the different factors. Oblique rotation was applied (0.05 divided by 6) so all effects are reported at a 0.008 level of significance. The effect size (r) for the difference between the control group and the HA group was −0.33, for the difference between the control group and the CI group was −0.29, and for the difference between the control group and the SSD group was −0.28. There were no significant differences between the hearing-impaired groups.

RESULTS

Figure 1 shows box plots of FAS and EAS scores for each group. The 50th percentile for FAS is around 14% for the control group but around 70% for each of the three hearing-impaired groups.

FAS Scores

Comparison of the FAS scores across all four groups revealed a significant difference (F[3] = 13.96, p < 0.05). Pairwise comparisons revealed a significant difference in FAS score between the control group (median FAS score = 13.75) and each of the individual hearing-impaired groups (HA group [median = 22.5]; CI group [median = 22.5]; SSD group [median = 22]; Controls [median = 22]). The effect size (r) for the difference between the control group and the HA group was −0.33, for the difference between the control group and the CI group was −0.29, and for the difference between the control group and the SSD group was −0.28. There were no significant differences between the hearing-impaired groups (HA versus CI: U = 1153, z = −0.67, p > 0.05; HA versus SSD: U = 1189, z = −0.42, p > 0.05; CI versus SSD: U = 1183, z = −0.46, p > 0.05).

The proportion of participants who experience extreme fatigue was 22, 10, and 22%, for the HA, CI, and SSD groups, respectively. The difference in the proportion of each group reporting extreme fatigue was not statistically significant (HA versus CI: χ²(1) = 2.68, p > 0.05; CI versus SSD: χ²(1) = 2.68, p > 0.05).

EAS Scores

Comparison of the EAS scores across all four groups revealed a significant difference (F[3] = 61.96, p < 0.05).

Table 3. EAS questions

<table>
<thead>
<tr>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do you have to put in a lot of effort to hear what is being said in conversation with others?</td>
</tr>
<tr>
<td>How much do you have to concentrate when listening to someone?</td>
</tr>
<tr>
<td>How easily can you ignore other sounds when trying to listen to something?</td>
</tr>
<tr>
<td>Do you have to put in a lot of effort to follow discussion in a class, a meeting, or a lecture?</td>
</tr>
<tr>
<td>Do you have to put in a lot of effort to listen on the telephone?</td>
</tr>
</tbody>
</table>

EAS, effort assessment scale.

Fig. 1. Boxplots of fatigue assessment scale and effort assessment scale scores. The solid horizontal line in the middle of each box plot represents the median score. Each box represents the upper and the lower quartiles of the data (the middle 50%). The distance between the upper quartile and the top whisker is the range of the top 25% scores. The distance between the lower quartile and the bottom whisker is the range of the bottom 25% scores. Circles represent outliers that are more than 1.5 times the interquartile range (the range between the upper and the lower quartile). Stars represent outliers that are more than three times the interquartile range.
Pair-wise comparison revealed a significant difference in EAS score between the control group (median EAS score = 20.2) and each of the individual hearing-impaired groups (HA group [median = 66.6]; U = 315.50, z = −6.37, p < 0.01, CI group [median = 70]; U = 359.00, z = −6.06, p < 0.01, SSD group [median = 70]; U = 250.00, z = −6.83, p < 0.01). The effect size (r) for the difference between the control group and the HA group was −0.64, for the difference between the control group and the CI group was −0.61, and for the difference between the control group and the SSD group was −0.69. There were no significant differences between hearing-impaired groups (HA control group and the SSD group was −0.69. There were no significant differences between hearing-impaired groups (HA versus CI: z = −0.18, p > 0.05; HA versus SSD: z = −6.83, p < 0.01, CI versus SSD: z = −6.06, p < 0.01). The effect size (r) for the difference between the control group and the HA group was −0.64, for the difference between the control group and the CI group was −0.61, and for the difference between the control group and the SSD group was −0.69. There were no significant differences between hearing-impaired groups (HA versus CI: z = −0.18, p > 0.05; HA versus SSD: z = −6.83, p < 0.01, CI versus SSD: z = −6.06, p < 0.01).

The proportion of participants who experience extreme listening effort was not statistically significant effort was 46, 54, and 52%, for the HA, CI, and SSD respectively. The difference in the proportion of each group reporting extreme listening effort was not statistically significant (HA versus CI: \( x^2(1) = 0.04, p > 0.05 \), HA versus SSD: \( x^2(1) = 0.36, p > 0.05 \), CI versus SSD: \( x^2(1) = 0.36, p > 0.05 \)).

**Correlation Between FAS and EAS Scores**

The scatter plot in Figure 2 shows that increased FAS scores are associated with increased EAS scores. There was a weak but significant correlation between the ratings of listening effort and fatigue of all four groups (r = 0.40, p < 0.05). There was a significant correlation between the EAS and the FAS scores for the CI group (r = 0.40, p < 0.05), the SSD group (r = 0.40, p < 0.05), and the control group (r = 0.30, p < 0.05). However, there was no significant correlation between FAS and EAS for the HA group (r = 0.20, p > 0.05).

We took the opportunity to investigate the relationship between the fatigue/effort score and: (a) the age and sex of participants and (b) severity of hearing loss in the HA group (indicated by the pure-tone average of each participant based on the results of a hearing test that was performed within 3 months of conducting the study). The correlation with the severity of hearing loss was not investigated for each of the CI and SSD groups. All of the participants in the CI group had bilateral profound hearing loss. All of the participants in the SSD group had one dead ear and passed a hearing screening at 30 dB HL in the other ear. There was no significant correlation between age and FAS scores (r = 0.021, p > 0.05) and between age and EAS scores (r = −0.042, p > 0.05) in any of the four groups. There was no significant difference in FAS and EAS scores between males and females (FAS: U = 4794.5, z = −0.230, p > 0.05, EAS: U = 4664.00, z = −0.553, p > 0.05). In the HA group, there was no correlation between the severity of hearing loss and FAS scores (r = −0.06, p > 0.05), and between the severity of hearing loss and EAS scores (r = 0.16, p > 0.05).

**Factor Analysis**

The KMO measure represents the ratio of the squared correlation between variables to the squared partial correlation between variables (Field 2009). The value of KMO ranges from 0 to 1. A value of 0 indicates that the sum of the partial correlation is large compared to the sum of correlations implying that there is a diffusion or a scatter in the pattern of correlation. A value of 0 means that performing a factor analysis is inappropriate as it will not be possible to identify distinct factors as a result of the scattered pattern of correlations. On the contrary, a value of 1 implies that the pattern of correlation is compact and that factor analysis will most probably yield distinct factors (Field 2009). In the present study, the KMO measure of 0.91 verified sampling adequacy for the analysis (Field 2009). The KMO statistics for individual variables was also satisfactory (above 0.5; Field 2009). Analysis yielded two factors with eigenvalues greater than one. The first factor had an eigenvalue of 6.56 and accounted for 41% of the variance. The second factor had an eigenvalue of 3.13 and accounted for 20.71% of the variance.

After rotation, the unique contribution of each variable to each factor is detailed in the pattern matrix in Table 4. Contributions less than 0.3 are not shown. The FAS questions (labeled as FAS 1 to FAS 10) load more strongly to the first factor (interpreted as fatigue) whereas the EAS questions (labeled EAS 1 to EAS 6) load more strongly to the second factor (interpreted as listening effort).

**TABLE 4. Pattern matrix: Unique contribution by each variable to each factor**

<table>
<thead>
<tr>
<th></th>
<th>Factor 1</th>
<th>Factor 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAS 1</td>
<td>0.705</td>
<td></td>
</tr>
<tr>
<td>FAS 2</td>
<td>0.763</td>
<td></td>
</tr>
<tr>
<td>FAS 3</td>
<td>0.488</td>
<td></td>
</tr>
<tr>
<td>FAS 4</td>
<td>0.568</td>
<td></td>
</tr>
<tr>
<td>FAS 5</td>
<td>0.738</td>
<td></td>
</tr>
<tr>
<td>FAS 6</td>
<td>0.798</td>
<td></td>
</tr>
<tr>
<td>FAS 7</td>
<td>0.653</td>
<td></td>
</tr>
<tr>
<td>FAS 8</td>
<td>0.846</td>
<td></td>
</tr>
<tr>
<td>FAS 9</td>
<td>0.736</td>
<td></td>
</tr>
<tr>
<td>FAS 10</td>
<td>0.507</td>
<td></td>
</tr>
<tr>
<td>EAS 1</td>
<td></td>
<td>−0.837</td>
</tr>
<tr>
<td>EAS 2</td>
<td></td>
<td>−0.847</td>
</tr>
<tr>
<td>EAS 3</td>
<td></td>
<td>−0.787</td>
</tr>
<tr>
<td>EAS 4</td>
<td></td>
<td>−0.881</td>
</tr>
<tr>
<td>EAS 5</td>
<td></td>
<td>−0.891</td>
</tr>
<tr>
<td>EAS 6</td>
<td></td>
<td>−0.789</td>
</tr>
</tbody>
</table>

EAS, effort assessment scale; FAS, fatigue assessment scale.

---

**Fig. 2.** Scatter plot of fatigue assessment scale and effort assessment scale scores for all participants and the linear regression line.
Psychometric Qualities of the EAS

Factor analysis indicated that all of the questions in the EAS load on to one factor which explained 71% of the variance. Factor loading for all of the EAS items was greater than 0.78. Intertitem correlation showed that the correlations between the different items of the EAS ranged from 0.63 to 0.83 and this provides confidence that all of the items in the EAS assess the same factor.

Internal consistency is the assessment of the degree of correlation between the different items of the scale (Bland & Altman 1997). Internal consistency of the EAS items was evaluated using Cronbach’s $\alpha$. For all items in the EAS this was 0.94. Removing any individual item of the EAS not improve Cronbach’s $\alpha$.

It was not possible to assess each of criterion validity and construct validity for the EAS in the present study. Criterion validity is “the extent to which a measure is empirically associated with relevant criterion variables” (Westen & Rosenthal 2003). Criterion validity depends on the presence of a “gold standard” that can be used in defining the concept of interest (Chrispin et al. 1997). The assessment of criterion validity was limited by the absence of a “gold standard” or specific criteria that can be used to confirm the hearing-impaired individual’s experience of listening effort. Construct validity is “the extent to which a measure adequately assesses the construct it purports to assess” (Westen & Rosenthal 2003). The assessment of construct validity is based on establishing the correlation between a potential measure and an established tool that is theoretically assessing the same construct (Westen & Rosenthal 2003). This is also not possible in the case of the EAS because of the absence of a standardized self-report measure for the assessment of self-reported listening effort.

In summary, there are five main findings in the study:

1. Hearing-impaired participants reported increased levels of listening effort and fatigue compared to the age-matched control group.
2. The proportion of participants who experience extreme fatigue was 22, 10, and 22%, for the HA, CI, and SSD groups, respectively. The proportion of participants who experience extreme listening effort was 46, 54, and 52%, for the HA, CI, and SSD groups, respectively.
3. There was no difference in self-reported levels of listening effort and fatigue between HA, CI, and SSD groups.
4. There was a weak positive correlation between FAS and EAS.
5. The questions of the FAS and the EAS assess two distinct constructs.

DISCUSSION

Self-Reported Fatigue

As hypothesized, hearing-impaired listeners reported increased levels of fatigue compared to the age-matched controls. This is consistent with the findings of Hornsby et al. (2014) who showed that hearing-impaired children also report greater levels of fatigue compared to age-matched controls. There was a wide range of fatigue scores in the HA group with around 10% of outliers. However, this was not the case with the other hearing loss groups who are significantly different from the controls. In any case, the difference between the HA group and the controls was still significant when the outliers were removed. It is apparent from Figure 1 that there was a wide range of fatigue scores and there is overlap between the hearing-impaired groups and the control group. The same pattern of overlapping results between hearing-impaired and control groups was also identified by Hornsby and Kipp (2016). All of the participants recruited in this study were elderly adults who may have been experiencing various levels of fatigue due to different reasons. However, it might be that for some people the addition of hearing impairment does not contribute significantly to the overall fatigue resulting in an overlap in FAS scores between the four groups.

The most likely explanation for the higher levels of self-reported fatigue in the hearing-impaired groups versus the control group is due to the hearing difficulty. Although it cannot be ruled out, we have no reason to believe that there were uncontrolled differences between the groups that could have systematically biased the results. It is possible that some chronic health conditions, for example, diabetes (Mitchell et al. 2009), may be more prevalent within the HA group versus the control group. However, differences in levels of chronic health conditions are unlikely to explain differences between the control group and the CI group (who have long-standing congenital hearing loss). With respect to the SSD group, participants might report increased levels of fatigue as a result of the surgery for removal of the acoustic neuroma (Ryeneman et al. 2004). However, this possibility was controlled for by only recruiting participants who had the surgery at least one year before the present study.

Results of the present study suggest that fatigue is a commonly reported problem in adults with different types of hearing impairment and there were no differences in fatigue between the hearing-impaired groups. It was expected that the difference in the severity of the hearing loss between the groups would correspond to differences in self-reported levels of fatigue between groups. For instance, it was expected that participants in the CI group might report the highest level of fatigue as a result of having profound hearing loss and that participants in the SSD might report least difficulties because of having one normal-hearing ear. The most likely explanation for the absence of a difference between the groups is the lack of correlation between self-reported fatigue and hearing level, which was also reported by Hornsby and Kipp (2016).

There are several explanations for the similarity in fatigue between the groups. First, it is possible that the participants in the hearing-impaired groups truly experience different levels of fatigue but the FAS is not sensitive enough to identify real differences between them. Second, participants in the hearing-impaired groups might experience similar levels of fatigue as a result of adjustments in their lifestyle to compensate for the hearing loss. Third, fatigue may be related to perceived hearing difficulty and not to the hearing level of the participants. In support of this last explanation, Hornsby and Kipp (2016) identified a correlation between self-reported fatigue and self-reported hearing difficulty but not between self-reported fatigue and the severity of hearing loss in hearing-impaired adults. Therefore, our future plans include investigating the correlation between perceived hearing difficulty and the FAS.

Smith et al. (2008) reported a mean raw FAS score of 16.5 and 15.3 in participants with chronic heart failure and stroke, respectively. By way of comparison, Smith et al. estimated the prevalence of extreme fatigue (based on scores above the
95th percentile of a control group on the FAS, as in the present study) for patients with stroke and chronic heart failure to be 61.3 and 67.3%, respectively. The raw median FAS score for the three hearing-impaired groups (9) is lower than FAS scores in patients with stroke and chronic heart failure. The higher prevalence of extreme levels of fatigue and higher mean FAS scores in patients with stroke and chronic heart failure reported by Smith et al. versus the levels reported by adults with hearing loss in the present study seems reasonable because fatigue is often the main symptom in stroke patients (Smith et al. 2008).

The FAS was not specifically developed for the assessment of fatigue in the hearing-impaired population. However, the findings of the present study suggest that the FAS may have potential to assess self-reported fatigue within people with hearing loss and compare levels of fatigue with other groups.

Self-Reported Listening Effort

Preliminary analysis of the psychometric qualities of the EAS indicated acceptable reliability and internal consistency of the EAS scale. It is commonly agreed that the minimum acceptable level of Cronbach’s α should be >0.7 (Field 2009), and for the EAS Cronbach’s α was 0.94. Confirming the validity of the EAS is limited by the lack of well-validated measures of listening effort against which the validity of the EAS might be established. Future work should consider the assessment of the content validity of the EAS, which would involve asking audiological experts and people with hearing impairment whether the items of the EAS are representative of the experience of listening effort.

Previous research has reported increased listening effort by hearing-impaired individuals for particular listening tasks in a research environment using scales like NASA-TLX (McCoy et al. 2005; Zekveld et al. 2010). Self-reported ratings of participants indicated that performing demanding listening tasks in the lab environment was perceived as effortful (Zekveld et al. 2010; Mackersie & Cones 2011). In the present study, hearing-impaired individuals reported significantly increased levels of listening effort in daily life compared to normal-hearing controls. The findings of the present study confirm that hearing-impaired individuals experience increased levels of listening effort not only in laboratory-based tasks but also in daily life. The findings of the present study also indicate that self-report measures of listening effort (e.g., EAS in this study) may be useful in indexing a potentially important aspect of hearing disability that is not indexed by traditional hearing assessment procedures. The findings are consistent with the hypothesis and with existing research (e.g., Kramer et al. 2006), that hearing-impaired individuals experience increased levels of listening effort in everyday life compared to normal-hearing controls. An important research need is to develop a validated scale that can be used in future studies for the assessment of listening effort in hearing-impaired individuals.

Based on the calculated effect size, we found that the magnitude of the difference in the self-reported listening effort between the control group and the hearing-impaired groups is larger than the difference in the self-reported fatigue. This might suggest that listening effort is more of a problem for the hearing-impaired population compared with fatigue. It is also possible that hearing-impaired participants are more aware of the experience of listening effort compared with fatigue.

Self-report measures have been previously used in audiology research to assess the effect of particular rehabilitation strategies on self-reported listening effort. For example, Noble and Gatehouse (2006) showed that HA users experienced lower levels of listening effort in daily life compared to hearing-impaired adults who were not HA users. This finding suggests that the use of hearing aids results in decreased self-reported listening effort. However, the findings of the present study suggest that the provision of hearing aids do not reduce listening effort to a level that is comparable to a normal-hearing individual. This should be interpreted with caution because we do not have information about participants’ daily use of the hearing aids. Self-report measures have also been used to assess listening effort in people with SSD. There is a common perception that SSD patients do not experience significant hearing difficulties as a result of having a normal-hearing ear (Douglas et al. 2007). Douglas et al. found that patients with SSD reported significant listening effort using the SSQ Hearing Scale compared to a matched control group, as in the present study. The findings of Douglas et al. (2007) and the present study suggests that individuals with SSD experience significant hearing disability due to increased listening effort.

In the present study, participants in the hearing-impaired groups reported approximately similar levels of listening effort despite differences in the severity of the hearing loss. The findings of the present study extend the findings of Hornsby and Kipp (2016) and suggest that hearing level might not be a valid predictor of self-reported listening effort. As discussed above in relation to fatigue, it is possible that the lack of a relationship between hearing level and self-reported listening effort is responsible for the lack of the difference in the self-reported listening effort between the three hearing-impaired groups. It is also possible that the EAS scale was not sensitive to differences in effort between the groups and that other measures such as measures of self-perceived hearing difficulty (Hornsby & Kipp 2016) might be more sensitive to differences between the groups. In addition, it is also possible that participants in the three groups may have modified their life styles to avoid particularly difficult listening situations and thus they do experience similar levels of listening effort.

Correlation Between Listening Effort and Fatigue

The weak but significant correlation between FAS and EAS supports the hypothesis that there is an association between effort and fatigue. Effort may lead to increased fatigue. However, it is also possible that high levels of fatigue may lead to increased levels of listening effort.

The weak correlation between FAS and EAS scores might be a result of the nature of the questions of the FAS and the EAS. The questions of the EAS are all hearing-specific whereas the questions of the FAS are general. It is possible that factors other than hearing disability influenced the EAS ratings. Participants in the present study were elderly adults who could have been experiencing high levels of fatigue as a result of a number of factors other than hearing difficulty, such as chronic illness. The contribution of multiple factors to the ratings of fatigue might have resulted in the weak correlation between FAS and EAS.

The possibility that the correlation between FAS and EAS reflect a general response bias rather than a real link between effort and fatigue also needs to be considered. In other words,
participants who provide low scores on one scale assessing one dimension may tend to provide low scores on other scales assessing other dimensions (Podsakoff & Organ 1986). It is difficult to control for response bias when investigating the relationship between any two self-report scales (Podsakoff & Organ 1986).

Factor Analysis

Factor analysis supported the hypothesis that the FAS and the EAS assess two distinct dimensions. Based on the content of the questions in the FAS and the EAS, we interpreted the factor that the FAS questions loaded onto as “fatigue” and the factor that the EAS questions loaded onto as “listening effort.” Future studies are required to determine the reliability and the sensitivity of FAS and EAS before they can be used as an outcome measure to compare interventions.

LIMITATIONS

Due to the cross-sectional design of this study, it was not possible to assess the effectiveness of hearing devices on listening effort and fatigue. This would require a controlled longitudinal study with assessment of listening effort and fatigue before and after intervention.

The range of hearing levels and age was limited. Recruiting participants with a wider range of hearing loss and age could facilitate a more thorough investigation of the correlation (or lack of) between the severity of hearing loss (and age) with listening effort and fatigue.

The hearing-impaired participants who participated in the present study may have been biased toward reporting increased levels of listening effort and fatigue, especially because the purpose of the study was not blinded, that is, participants were aware that effort and fatigue was being investigated in individuals with a hearing loss. The possibility that the participants have been biased toward reporting high level of listening effort highlights the potential importance of identifying a physiological measure of effort and fatigue to be used alongside self-report measures. Using a combination of self-report and physiological measures may help elucidate the factors that contribute to self-ratings of effort and fatigue. This recommendation assumes that there is a relationship between processing effort and perceived effort and this may not be the case.

The weak correlation identified between the FAS and the EAS suggest that other variables such as self-perceived hearing difficulty, the presence of chronic health conditions, and the lifestyles of hearing-impaired individuals should be investigated as predictors of listening effort and fatigue.

Finally, the present study focused on the adult population and it would be helpful to investigate effort and fatigue in other age groups to identify whether our findings apply to other age groups of hearing-impaired individuals.

CONCLUSIONS

The main conclusions are

1. Hearing-impaired adults report high levels of listening effort and fatigue in their daily life.
2. 2 out of 10 participants reported extreme levels of fatigue and 5 out of 10 participants reported extreme levels of effort.
3. Adult HA, CI, and those with SSD reported similar levels of effort and fatigue suggesting that these cannot be predicted from hearing level.
4. The FAS and the EAS assess two distinct dimensions.

ACKNOWLEDGMENTS

The authors thank the Manchester Biomedical Research Centre and the Greater Manchester Comprehensive Local Research Network. The authors also thank Andrea Wadeson, Sinead Toal, Unai Martinez de Estibariz, Deborah Mawman, Nathan O’Doherty, The University of the Third Age, and The Irish Community Care for their help in recruiting participants.

The authors have no conflict of interest to disclose.

Address for correspondence: Sara Alhanbali, Manchester Centre for Audiology and Deafness, University of Manchester, Manchester M13 9PL, United Kingdom. E-mail: sara.al-hanbali@postgrad.manchester.ac.uk

Received September 29, 2015; accepted June 24, 2016.

REFERENCES


**REFERENCE NOTE**

Addendum to Study One

The following five points were based on comments and discussion with the examiners at the PhD viva on 14 December 2017.

1- The reports of increased listening effort and fatigue might be due to participants’ perception of their hearing disability. Participants might have assumed that having a hearing problem would intuitively suggest that they experience increased levels of listening effort and fatigue even if they do not actually experience these listening difficulties in everyday life.

2- Approaching participants who had their hearing tested within 3 months of completing the questionnaires might have biased the results towards reporting increased levels of listening effort because these individuals were being seen because of on-going hearing difficulties. An alternative approach would have been to test the hearing level of each participant. Asking participants to have their hearing tested before completing the questionnaires might have also biased the results if only certain people agreed to attend for this additional testing.

3- Participants in the control group were approached through social groups. It is possible that these adults are more energetic and generally have better health condition compared to participants in the other groups. Therefore, approaching participants through social groups might act as a confound for the finding that controls reported less fatigue. An alternative approach would have been to recruit hospital patients attending non-audiology appointments. However, recruiting hospital patients might have increased the chances of controls having chronic health conditions that might increase levels of fatigue.
4- Page e43, paragraph “FAS Scores”: typos in the p values lines 6, 7, and 8. All p values should be 0.008.

Page e44, paragraph “EAS Scores”: typos in the p values lines 6, 7, and 8. All p values should be 0.008.

5- There was a non-significant trend of increased self-reported effort with increased hearing levels. The correlation between hearing level and EAS ($r=0.16$) was based on the hearing levels of participants in the HA group only. Hearing thresholds for the participants in the control group were not available as they all passed a hearing screening at 30 dB HL. Including a wider range of hearing levels in the analysis might have resulted in a significant correlation between hearing level and EAS. The significant difference in self-reported listening effort and fatigue between the control group and the groups with hearing impairment suggest that it might not be possible to reject the hypothesis that the severity of hearing impairment is related to listening effort and fatigue.
CHAPTER FOUR

STUDY TWO: HEARING HANDICAP AND SPEECH RECOGNITION CORRELATE WITH SELF-REPORTED LISTENING EFFORT AND FATIGUE

This manuscript has been accepted for publication in *Ear and Hearing*:


The format of *Ear and Hearing* manuscripts is used in the chapter.

Page number of thesis: 79

---

1 Self-report scales used in Study Two are provided in Appendix B (Fatigue Assessment Scale), Appendix C (Effort Assessment Scale), and Appendix F (Hearing Handicap Inventory for Elderly)
Hearing Handicap and Speech Recognition Correlate With Self-Reported Listening Effort and Fatigue

Sara Alhanbali,1,3 Piers Dawes,1,3 Simon Lloyd,2,3 and Kevin J Munro1,3

Objectives: To investigate the correlations between hearing handicap, speech recognition, listening effort, and fatigue.

Design: Eighty-four adults with hearing loss (65 to 85 years) completed three self-report questionnaires: the Fatigue Assessment Scale, the Effort Assessment Scale, and the Hearing Handicap Inventory for Elderly. Audiometric assessment included pure-tone audiometry and speech recognition in noise.

Results: There was a significant positive correlation between handicap and fatigue (r = 0.39, p < 0.05) and handicap and effort (r = 0.73, p < 0.05). There were significant (but lower) correlations between speech recognition and fatigue (r = 0.22, p < 0.05) or effort (r = 0.32, p < 0.05). There was no significant correlation between hearing level and fatigue or effort.

Conclusions: Hearing handicap and speech recognition both correlate with self-reported listening effort and fatigue, which is consistent with a model of listening effort and fatigue where perceived difficulty is related to sustained effort and fatigue for unrewarding tasks over which the listener has low control. A clinical implication is that encouraging clients to recognize and focus on the pleasure and positive experiences of listening may result in greater satisfaction and benefit from hearing aid use.

Key words: Fatigue, Listening effort.

INTRODUCTION

Alhanbali et al. (2017) reported higher levels of self-reported listening effort and fatigue in adults with hearing loss compared with controls with good hearing. Consistent with recent studies (e.g., Petersen et al. 2015; Hornsby & Kipp 2016), there was no correlation between hearing level and listening effort or fatigue. The lack of correlation suggests that audibility, per se, is not the cause of listening effort or fatigue.

Hornsby and Kipp (2016) reported a positive correlation between self-reported fatigue and hearing difficulty (using the Hearing Handicap Inventory for Elderly [HHIE, Ventry & Weinstein 1982] or Adults [Newman et al. 1990]). The finding of Hornsby and Kipp is consistent with the Motivation Control Model (MCM) of effort and fatigue proposed by Hockey (2013). This model views fatigue as an adaptive state that maintains efficient prioritization and management of competing tasks. The subjective experience of fatigue arises when there is conflict between current and alternative tasks. If a demanding task, over which an individual has little control, is perceived as resulting in low success, the individual experiences fatigue. As a result, individuals may modify their behavior and reduce effort on the demanding task (i.e., avoid fatigue at the expense of reduced task performance) or prioritize a task that is less demanding or more rewarding.

Hockey’s MCM describes a triangular relationship between aspects of control, task demands, and perceived reward. Fatigue is likely in demanding conditions over which individuals have little control if increased effort is not perceived as resulting in successful performance. The following example shows how the model might relate to having a hearing problem. An individual is motivated to listen to conversation in a noisy party. As the level of background noise (i.e., the demands) increases, the individual might continue to exert listening effort to prevent deterioration in performance. However, the individual might get to a point where he is unable to carry on because the background noise (outside their control) is too loud. In this case, the individual might lose motivation because the perceived rewards of the task are not sufficient, and sustained effort results in fatigue.

Hockey’s MCM suggests that fatigue results from sustained effort in situations perceived as unrewarding. This is consistent with the Framework for Understanding Effortful Listening (Pichora-Fuller et al. [2016]). According to the Framework for Understanding Effortful Listening, motivation to engage in task performance is also likely to result in increased listening effort when performance is perceived as rewarding.

Hockey’s MCM describes transient states of fatigue that occur as a result of experiencing periods of sustained effort (Hockey 2013; Hornsby & Kipp 2016). Adults with hearing loss likely experience periods of sustained effort in daily listening situations. Chronic fatigue may occur if there is sustained effort with little opportunity for recovery. The mental stress associated with having to communicate might persist even when individuals are not involved in a listening task.

AIMS

To our knowledge, no studies have been designed specifically to investigate the correlation between (1) hearing handicap (disability and handicap now called “activity limitation” and “participation restrictions,” respectively, in the International Classification of Functioning Disability and Health; World Health Organisation 2001) and listening effort, or (2) speech recognition and listening effort or fatigue. The aims of this study were (1) to investigate the correlation between hearing handicap and both self-reported listening effort and fatigue, and (2) to investigate the correlation between speech recognition and both self-reported listening effort and fatigue. A wide range of behavioral and physiological measures exist for the assessment of listening effort and fatigue. Examples of behavioral measures that have been used include reaction time, for example, Houben et al. (2013) and dual task, for example, Desjardins and Doherty.
(2013). Examples of physiological measures include galvanic skin response, for example, Macksery et al. (2015), electroencephalography, for example, Petersen et al. (2015), and pupillometry, for example, Zekveld et al. (2011). The focus of the present article was on self-report measures due to (1) the ease of administration and (2) to facilitate comparability with Hornsby and Kipp (2016).

Despite exerting considerable effort, individuals with hearing loss report difficulties in challenging listening situations (Hornsby & Kipp 2016). According to Hockey’s MCM, sustained listening effort, with low reward, leads to fatigue. We hypothesized that hearing handicap would have a stronger correlation with self-reported listening effort and fatigue than lab-based speech recognition, because the former assesses hearing problems in real life situations.

METHODS

Participants
Eighty-four adults with hearing loss (mean age 72 years, SD: 6, range 65 to 85) were identified through the database of a UK National Health Service audiology department (Withington Community Hospital, Manchester). A minimum sample size of 77 was estimated to provide 80% statistical power with a medium effect size ($f^2 = 0.15$), according to a Cohen $f^2$ effect size method. All participants were native English speakers and had bilateral sensorineural hearing loss ranging from mild to severe based on the better ear pure-tone average (PTA) hearing threshold level at 0.5, 1, 2, and 4 kHz. According to the British Society of Audiology’s, recommended procedures, a hearing threshold in the range of 20 to 40 dB HL is classified as a mild hearing loss, and a hearing threshold in the range of 71 to 95 dB HL is classified as a severe hearing loss. Better ear 4-frequency average was 47 dB HL (SD: 15.4, range 25 to 80). Participants who were hearing aid users used their devices most of the day for a period of at least 6 months. Participants’ daily hearing aid use was evaluated based on their response to the question “Do you use your hearing aid most of the day?” with response options of yes or no. Of the 84 participants, eight were not hearing aid users, 26 were unilateral hearing aid users, and 50 were bilateral hearing aid users. Participants with a diagnosed cognitive condition, such as dementia, were not recruited.

Outcome Measures
Following Alhanbali et al. (2017), fatigue was assessed using the 10-question Fatigue Assessment Scale (FAS, Michielsen et al. 2004, see supplemental file/Appendix 1, Supplemental Digital Content 1, link: links.lww.com/EANDH/A388). The FAS is a standardized generic scale of fatigue with good internal consistency, reliability, and validity (Michielsen et al. 2004). Participants were asked to rate how they feel on a daily basis, for example, I get tired very quickly (3-point Likert scale from never to always). Effort was assessed using the six-question Effort Assessment Scale (EAS), developed by Alhanbali et al. (2017), which includes the three effort-related questions from the Speech, Spatial Quality (SSQ) Hearing Scale (Gatehouse & Noble 2004), for example, “How much do you have to concentrate when listening to someone?” (10-point visual analogue scale; 0 is no effort and 10 is lots of effort, see supplemental file/Appendix 2, Supplemental Digital Content 2, link: links.lww.com/EANDH/A389). The EAS is not a standardized scale; however, results of Alhanbali et al. (2017) suggested that the scale has good internal consistency and that all of its items load into a single dimension. The 25-question HHIE was used to assess perceived hearing difficulties so that the findings could be compared with Hornsby and Kipp (2016) for example, “Does a hearing problem cause you to avoid groups of people?” (three response options: “yes,” “sometimes,” or “no”). The HHIE is a standardized scale that has good internal consistency and test–retest reliability (Ventry & Weinstein 1982).

Analysis
Speech recognition in noise was measured using an adaptive procedure to determine the signal to noise ratio (SNR), in decibels, required for 71% correct. The testing was completed in a sound-treated booth. Participants performed the speech test with their hearing aids on. Participants performed the listening task at the normal user setting. Participants verified verbally that their hearing aids were functioning adequately and that they were able to clearly hear the voice of the researcher at a normal conversational level. The speech material used was a monosyllabic presentation of the digits “1” to “9” (excluding the bisyllabic “7”). The digits were taken from the conversational speech level recordings in the Whispered Voice Test (McSheffery et al. 2013). Digits in noise test was used because it (1) is a widely used task, and (2) yields comparable data across different language groups (Dimitrijevic et al. 2017). Strings of three digits, spoken by a male speaker, were presented from two loudspeakers placed 1 m away from where the participant sits at ±45° azimuth. The digits were presented at a level of 65 dBA in the presence of stationary background noise that started 5 sec before the first digit and ended 1 sec after the end of the last digit. Presenting 5 sec of noise before the start of the first digit was expected to be sufficiently long for the noise reduction algorithm to be activated in the hearing aids.

Participants were presented with groups of three digits at a time. After the presentation of each group of three digits, a box appeared on the screen positioned 50 cm from the participant. Participants responded by clicking on the numbers they heard using a computer mouse. A response was only considered correct if the participant correctly identified all three digits and in the correct order. In the first 10 presentations, the level of noise increased by 3 dB in the case of a correct response and decreased by 3 dB in the case of an incorrect response. In subsequent reversals, the level of noise varied in a 2 dB down 1 dB up adaptive procedure until the 71% correct performance level was established. The SNRs of both of the last reversal (trial) and the mean of the last 10 reversals were recorded. We inadvertently used the SNR of the last reversal in the main experiment; however, we are reassured by the similarity between the two methods (<1 dB difference) that speech was presented at around 71% for each participant. The SNR was calculated based on a single test administration. The duration of the listening task ranged from 12 to 15 min depending on the response time of each participant. Testing was completed in a single session.
The correlations between hearing handicap, SNR, hearing level, fatigue, and effort were analyzed using Spearman correlation coefficient. Forced entry multiple linear regression was used in which FAS or EAS was the dependent variable and the HHIE, age, and PTA were the independent variables. Forced entry multiple linear regression was also carried out to investigate the effect of age and PTA on the correlation between (1) FAS and SNR, (2) EAS and SNR, and (3) FAS and EAS.

Visual examination of scatter plots suggested that a linear model could provide the best representation for all of the aforementioned analyses, and this was confirmed with the curve estimation function on SPSS (IBM statistics SPSS version 22) and the norm of residuals function on MATLAB (MathWorks, version 2015a).

**RESULTS**

Median and interquartile range (IQR) for the different self-report scales were as follows: HHIE: median = 38.00, IQR = 36.00; FAS: median = 20.00, IQR = 17.50; EAS: median = 71.67, IQR = 31.67. Based on the reference data from Ventry and Weinstein (1982), 41.86% of the participants had significant handicap (HHIE scores greater than or equal to 43), 41.86% of the participants had mild to moderate handicap (HHIE scores between 17 and 42), and 16.28% had no handicap (HHIE scores less that 17). Figure 1 shows scatter plots of the relationship between the different variables. Spearman ρ and the significance values are provided on each plot. There was a statistically significant positive correlation between hearing handicap and fatigue and also effort. There were no significant correlations between hearing level and fatigue or effort. There was no significant correlation between age and both hearing handicap and fatigue. The was a weak significant correlation between age and effort. Correlations between fatigue/effort and hearing handicap remained unchanged in multiple regression models that included age and hearing level with the HHIE being the only significant predictors in both of the models (Table 1).

There was a significant positive correlation between worse speech recognition and greater effort and also greater fatigue, that is, the need for a more positive SNR was associated with greater effort/fatigue. Correlations between effort and speech recognition remained unchanged in multiple regression models that included significant predictors of age and hearing level. The correlation between fatigue and speech recognition became insignificant in multiple regression models that included age and hearing level, suggesting that these factors might have an influence on the correlation between fatigue and speech recognition (Table 1).

There was a significant positive correlation between effort and fatigue. Correlations between fatigue and listening effort remained unchanged in multiple regression models that included age and hearing level with the effort being the only
significant predictors of fatigue (Table 1). However, the weak value of $R^2$ suggests that the correlation between fatigue and listening effort is of minimal significance, and other variables are likely to influence participants’ experience of fatigue.

**DISCUSSION**

Consistent with the findings of Hornsby and Kipp (2016), there was (1) a significant correlation between self-reported fatigue and hearing handicap, but (2) no correlation between self-reported fatigue and hearing level. Our findings also show a similar pattern for self-reported listening effort, which is consistent with the findings of Eckert et al. (2017). When developing the SSQ Hearing Scale, Gatehouse and Noble (2004) also identified a correlation between hearing handicap and effort. The same pattern was also reported when using the SSQ Hearing Scale to investigate the effect of interaural asymmetry of hearing loss (Noble & Gatehouse 2004) and the effect of using one versus two hearing aids (Noble & Gatehouse 2006). We have also identified significant but weaker correlations between self-reported listening effort/fatigue and lab-based speech recognition.

Our findings are consistent with Hockey’s MCM where fatigue is a control mechanism to limit investment of resources in an unrewarding activity over which the listener has little control. Sustained effort is a precursor to fatigue but, according to Hockey’s MCM, fatigue is not a direct consequence per se. Fatigue is a consequence of increased effort when performance is not perceived as rewarding. This link between sustained effort and fatigue may explain the correlations between self-reported listening effort and fatigue observed in the present study. Given the possible causal relationship between the experiences of listening effort and fatigue, this could explain why perceived hearing handicap is correlated with both listening effort and fatigue. The correlations with fatigue may have been stronger if the fatigue measure had focused specifically on listening instead of general fatigue. It is also possible that factors such as the age of the participants may have influenced self-reported fatigue. Results of regression analysis suggested that the correlation between hearing handicap and self-reported fatigue was independent of the effect of age and hearing sensitivity. However, considering that fatigue is likely to increase with age (Avlund 2010), the absence of an effect of age might be due to the limited age range of the participants recruited in this study (65 to 85 years).

There was a significant (but weaker) correlation between speech recognition and both listening effort and fatigue. On average, listeners with poorer speech recognition (i.e., those who require a more positive SNR) reported greater listening effort compared with listeners who achieved criterion performance at a more negative SNR. Our findings are consistent with Eckert et al. (2017) who reported a correlation between perceived listening effort and performance on a sentence in noise task. There was considerable variability in the scores, suggesting that it would not be ideal to use performance on the speech test to predict listening effort for a given individual. The weak relationship between lab-based measures of speech recognition and effort/fatigue is unsurprising given the lack of correlation between PTA and effort/fatigue in the present study and in previous research (e.g., Hornsby & Kipp 2016; Alhanbali et al. 2017). Speech recognition in quiet, and to some extent in noise, is correlated with hearing thresholds (Vlaming et al. 2014). Detection of pure tones and speech recognition do not necessarily reflect individual differences in hearing handicap. Self-report and performance-based measures may assess different aspects of the same experience (Pichora-Fuller et al. 2016). Factors such as motivation or boredom might influence performing listening tasks in the lab therefore weakening the correlations with speech recognition versus those involving hearing handicap. Although performing the speech in noise task took about 10 min, participants may have been bored due to the repetitive nature of the task despite its short duration (Hockey 2013).

Our findings show that perceived communicative success (indicated by hearing handicap) and listening effort and fatigue are related. Listening effort and fatigue can have a negative impact on quality of life and limits the benefit of hearing aids (McGarrigle et al. 2014; Pichora-Fuller et al. 2016). Therefore, it may be useful to measure listening effort and fatigue to facilitate optimal hearing care. Measuring listening effort and fatigue will provide a more comprehensive assessment of hearing disability. In addition, it may be possible to use effort/fatigue as an outcome measure when providing intervention (or comparing different interventions, e.g., amplification with noise reduction enabled/disabled, Pichora-Fuller et al. [2016]).
Based on the theory of rational motivational arousal, Matthen (2016) argues that motivation and pleasure have a major role in alleviating the negative experiences associated with hearing loss. A hearing rehabilitation strategy that improves audibility and focuses on successful task performance is less likely to alleviate negative emotions, such as displeasure and fatigue. Identifying ways of encouraging the client to recognize and focus on the pleasure and positive experiences of listening, even when demanding, may be beneficial. Focusing on the positive experiences of listening is likely to improve patient’s satisfaction and improve hearing aid use. There is the potential for audiologists to collaborate with health psychology to develop ways of achieving this goal. Educating patients on how to minimize the demands they encounter by selecting and modifying different listening situations whenever possible might be also beneficial. The effect of altering aspects of task demand and motivations on the experience of listening should be emphasized in the rehabilitation process (Pichora-Fuller et al. 2016). Pichora-Fuller et al. (2016) identified factors, such as stress, stigma, and low self-efficacy (which may be related to motivational factors and reward), as having a negative influence on the performance of individuals with hearing loss in everyday listening situations. They recommend a consideration of social and psychological factors in aural rehabilitation to boost hearing aid benefit.

A limitation of our correlational design is that the direction of any causal relationship cannot be established. Another potential limitation is the correlation between lab-based speech-in-noise measures and daily life measures of effort/fatigue. A stronger correlation might have been identified if the assessment was restricted to self-reported effort/fatigue to the lab-based task. The correlation between handicap and effort (0.78) is stronger than the correlation between handicap and fatigue (0.39). The difference in the size of the correlation could be due to differences in sensitivity of the effort and fatigue questionnaires because the effort questions are hearing specific and the fatigue questions probe general experience. Alternatively, sustained effort will not always lead to fatigue if, for example, the listener is motivated to engage in the task (“I want to do this task” instead of “I have to do this task”) because it is under the control of the listener and performance is perceived as rewarding.

CONCLUSIONS

Self-reported listening effort and fatigue are positively correlated with hearing handicap and lab-based measures of hearing difficulty but not hearing level. This is consistent with the Motivational Control Model where perceived difficulty is related to sustained effort and fatigue for unrewarding tasks over which the listener has low control. To our knowledge, we are the first to show a correlation between (1) hearing handicap and listening effort, and (2) speech recognition and listening effort and fatigue in a study that was specifically designed to investigate these correlations. The correlations with lab-based measures of performance are lower than for handicap and suggest that actual performance is affected by multiple factors.

ACKNOWLEDGMENTS

This Manchester Biomedical Research Centre is funded by the National Institute for Health Research.

The authors have no conflicts of interest to disclose.

Address for correspondence: Sara Alhanbali, Manchester Centre for Audiology and Deafness, School of Health Sciences, University of Manchester, Manchester, M13 9PL, UK. E-mail: sara.al-hanbali@postgrad.manchester.ac.uk

Received March 7, 2017; accepted September 13, 2017.

REFERENCES


Addendum to Study Two

The following four points were based on comments and discussion with the examiners at the PhD viva on 14 December 2017.

1- There was a missing number in the scale of PTA. Revised figure:

![Scatter plots showing age, pure-tone average (PTA), signal-to-noise ratio (SNR), Hearing Handicap Inventory for Elderly (HHIE), Fatigue Assessment Scale (FAS), and Effort Assessment Scale (EAS) scores for all participants (n = 84). Spearman correlation coefficient is provided on each scatter plot. **Correlation is significant at the 0.01 level (2 tailed). *Correlation is significant at the 0.05 level (2 tailed).](image)

Figure 1. Scatter plots showing age, pure-tone average (PTA), signal-to-noise ratio (SNR), Hearing Handicap Inventory for Elderly (HHIE), Fatigue Assessment Scale (FAS), and Effort Assessment Scale (EAS) scores for all participants (n = 84). Spearman correlation coefficient is provided on each scatter plot. **Correlation is significant at the 0.01 level (2 tailed). *Correlation is significant at the 0.05 level (2 tailed).

Figure 1 demonstrates a non-significant trend of increased self-reported effort with increased severity of hearing impairment. The number of the participants included when investigating the correlation between EAS and PTA is less than the number of participants included when investigating the correlation between FAS and SNR due to removing a single outlier. The relationship between EAS and PTA was significant before removing the
outlier (rs=0.24, p= 0.03) This was not included in the figure because of the disproportionate effect of a single participant.

2- The findings of Study Two are consistent with Hockey’s Motivational Control Model (2013). However, it is important to be cautious about the hypothesised link because the model describes transient states of fatigue while the manuscript describes long-term fatigue. The model was assumed to generalise to individuals with long-term fatigue because: i) long term fatigue is a consequence of short term fatigue that is not followed by long enough recovery periods, ii) the mental stress associated with having to communicate might persist even when individuals are not involved in a listening task.

3- Hockey’s model suggests that fatigue is usually a consequence of increased levels of effort. However, it is important to note that the model also suggest that increased fatigue might be associated with decreased effort as a result of giving up on task performance.
CHAPTER FIVE

STUDY THREE: Is Listening Effort Multidimensional?2

This chapter is currently under review in *Ear and Hearing*.

The manuscript format of *Ear and Hearing* is used in the chapter.

---

2 The procedure for transferring pupil size from pixels to mm is provided in Appendix K
Simultaneous recording of multi-modal measures demonstrate that listening effort is multidimensional

*Sara Alhanbali*\(^1,2\), *Piers Dawes*\(^1,2\), *Rebecca E Millman*\(^1,2\), and *Kevin J Munro*\(^1,2\)

\(^1\) Manchester Centre for Audiology and Deafness, School of Health Sciences, University of Manchester, Manchester, M13 9PL

\(^2\) NIHR Manchester Biomedical Research Centre, Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, M13 9WL

**Financial Disclosures/Conflicts of Interest:** None

Address correspondence to: Sara Alhanbali, Manchester Centre for Audiology and Deafness, School of Health Sciences, University of Manchester, Manchester, M13 9PL, UK.

Phone: +44 161 275 8568

E-mail: sara.al-hanbali@postgrad.manchester.ac.uk
Abstract

Objective: The literature on listening effort is as confusing as it is voluminous: measures of listening effort rarely correlate with each other and sometimes result in contradictory findings. Here, for the first time, we directly compared simultaneously recorded multi-modal measures of listening effort. After establishing the reliability of the measures, we investigated validity by quantifying correlations between measures and then grouping related measures through Factor Analysis.

Design: 116 participants with hearing levels ranging from normal to severe took part in the study (age range: 55-85 years old, 50.3% male). Listening effort was measured simultaneously using reaction time, pupil size, electroencephalographic alpha power, skin conductance, and a self-report measure. One self-report measure of fatigue was also included. The listening task involved correct recall of a random digit from a sequence of six presented at a signal-to-noise ratio where criterion performance was around 71%. Test-retest reliability of the measures was established by re-testing 30 participants 7 days after the initial session.

Results: With the exception of skin conductance and the self-report measure of fatigue, interclass correlation coefficients (ICC) revealed good test-retest reliability (minimum ICC: 0.75). Weak or non-significant correlations were identified between measures. Factor Analysis, using only the reliable measures, revealed four underlying dimensions: Factor 1 included SNR, hearing level, performance accuracy, and baseline alpha power; Factor 2 included pupillometry; Factor 3 included alpha power (at baseline, during speech presentation and during retention) and self-reported listening effort; Factor 4 included reaction time, self-reported listening effort, and baseline alpha power.
**Conclusion**: The good ICC suggests that poor test reliability is not the reason for the lack of correlation between measures. For the first time, we have demonstrated that the measures that have been traditionally used as indicators of listening effort tap into multiple underlying dimensions that we interpret as: performance (Factor 1), task engagement (Factor 2), cognitive processing (Factor 3), and behavioural consequences (Factor 4). The underlying dimensions assessed by the different measures might not be necessarily related to listening effort. This suggests that measures should not be used interchangeably. This finding also provides a framework for understanding and interpreting listening effort measures, and has widespread implications for both research and clinical practice.
Introduction

Pichora-Fuller et al. (2016) have recently defined listening effort as “the deliberate allocation of mental resources to overcome obstacles in goal pursuit when carrying out a task”. It has been traditionally assumed that the experience of listening effort is predominantly influenced by the demands of the listening task. However, recent interpretations of the concept of listening effort and its underlying mechanisms suggest that multiple dimensions influence the experience of listening effort (e.g. Pichora-Fuller et al. 2016; Peelle 2017; Strauss and Francis 2017). The deliberate allocation of cognitive resources required to justify sustained effort is influenced by the motivation and reward associated with perceived performance. Therefore, Strauss and Francis (2017) suggested that in demanding listening tasks, it is not possible to assume that the amount of effort required to complete the task (demanded effort) equals the amount of effort that individuals actually exert (exerted effort). Individual variability in behaviorally or physiologically measured listening effort on the same task can result from differences in factors such as motivation and arousal. The influence of multiple factors on the experience of listening effort suggests that listening effort might be a multidimensional process. In support for the multidimensionality of listening effort, Peelle (2017) suggests that multiple cognitive systems are activated during effortful listening.

Individuals with hearing impairment report increased listening effort in everyday life despite using hearing aids or cochlear implants (Alhanbali et al. 2017a). People who experience increased listening effort are likely to report increased negative impacts on the social and emotional aspects of their life (Alhanbali et al. 2017b). Sustained listening effort is thought to result in the development of listening-related fatigue in situations
where the increased effort is not perceived as resulting in successful performance (Hockey 2013; Alhanbali et al. 2017b). Listening-related fatigue has been defined as “extreme tiredness resulting from effortful listening” (McGarrigle et al. 2014). Identifying reliable clinical measures of listening effort may provide a means of indexing an important dimension of hearing disability that is currently not well captured by current audiological measures such as pure-tone and speech audiometry, or self-reported measures of disability or handicap (disability and handicap now called “activity limitation” and “participation restrictions”, respectively, in the International Classification of Functioning Disability and Health; World Health Organisation 2001). A clinical measure of listening effort could also inform interventions that redress these important aspects of hearing disability.

In research settings, various purported measures of listening effort have been used including: i) self-report such as NASA Task Load Index (Hart and Staveland 1988), ii) behavioural such as reaction time e.g. Houben et al. (2013) and dual task e.g. Desjardins and Doherty (2013), and iii) physiological including galvanic skin response e.g. Mackersie et al. (2015), electroencephalographic measures e.g. Petersen et al. (2015) and pupillometric indices e.g. Zekveld et al. (2011). However, it is not clear if these measures tap into the same construct and this may explain, at least in part, why the different measures rarely correlate with each other (McGarrigle et al. 2014). Multiple measures of listening effort have generally not been obtained simultaneously while the participant performs a listening task, making it difficult to make a direct comparison between the measures. The reliability of alternative listening effort measures must be established before they could be considered for use in research or clinical settings (Koo and Li 2016).
Unreliable measures are unlikely to correlate strongly with each other, even if they index the same construct.

**Measures of listening effort and fatigue**

McGarrigle et al. (2014) and Pichora-Fuller et al. (2016) provide a detailed discussion of the self-report, behavioural, and physiological measures that have been used in listening effort/fatigue research. Ohlenforst et al. (2017a) also provides a systematic review of studies that investigated the effect of hearing impairment or the effect of hearing aid amplification on listening effort. Table 1 provides a summary of the measures and their main advantages and disadvantages.
Table 1. The advantages and disadvantages of using self-report, behavioural, and physiological measures of listening effort.

<table>
<thead>
<tr>
<th>Measures</th>
<th>Advantages</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-report</td>
<td>• Quick and easy to administer</td>
<td>• Affected by individual differences in interpreting questionnaires</td>
</tr>
<tr>
<td></td>
<td>• In everyday life, e.g. Alhanbali et al. (2017a)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• In research settings, e.g. Mackersie and Cones (2011)</td>
<td></td>
</tr>
<tr>
<td>Behavioural</td>
<td>• Easy to administer and interpret</td>
<td>• Can be affected by individual differences in aspects such as motivation and task engagement</td>
</tr>
<tr>
<td></td>
<td>• Reaction time, e.g. Houben et al. (2013)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Dual task, e.g. Desjardins and Doherty (2013), Sarampalis et al. (2009)</td>
<td></td>
</tr>
<tr>
<td>Physiological</td>
<td>• Provides precise temporal indications about mental processing</td>
<td>• Difficult to discriminate between good effort (associated with improved performance) and bad effort (reflecting strain to cope with increased task demands) e.g. pupillometry (Ohlenforst et al. 2017b)</td>
</tr>
<tr>
<td></td>
<td>• Pupillometry, e.g. Zekveld et al. (2010)</td>
<td>• Consistency of the findings is affected by how demanding the task is (e.g. Obleser and Weisz 2012 and McMahon et al. 2016)</td>
</tr>
<tr>
<td></td>
<td>• EEG, e.g. Obleser and Kotz (2011)</td>
<td>• Can be affected by individual differences in aspects such as motivation and task engagement (Wendt et al. 2016)</td>
</tr>
<tr>
<td></td>
<td>• Skin conductance, e.g. Mackersie et al. (2015)</td>
<td></td>
</tr>
</tbody>
</table>

The literature on listening effort is as confusing as it is voluminous. Inconsistencies between different measures of listening effort and fatigue have been reported including disagreement between different: (i) measures, (ii) participant groups, and (iii) studies that used the same measure to test similar groups of participants but used different listening tasks. The variability in the testing methods used across the different studies (including
speech material, participants, listening conditions) complicates the ability to directly compare their results (Ohlenforst et al. 2017a). Therefore, it is not clear if these inconsistencies are because the measures assess different processes, or because some measures are unreliable or lack sensitivity. Further discussion on the inconsistencies reported in the literature is provided in the following section.

Disagreement between measures

Several authors have reported no correlation between self-report and behavioural/physiological measures of listening effort or fatigue. For instance, Wendt et al. (2016) compared listening effort in participants with high and low working memory capacity and reported contradictory findings for pupillometric and self-report measures of listening effort. Increased arousal indexed by larger pupil sizes during a listening task has previously been interpreted as reflecting increased effort, e.g. Zekveld et al. (2011). However, Wendt and colleagues reported that participants with high working memory capacity reported less effort than participants with low working memory capacity yet they had larger pupil dilation. Wendt et al. suggested that higher levels of attentional focus and vigilance reflected by larger pupil dilation do not necessarily translate to perceived effort and strain. The findings of Wendt et al. are consistent with the idea that self-report and pupillometric measures of listening effort may actually index different constructs. In Wendt and colleagues’ work, pupillometry was sensitive to increased cognitive demands that resulted in perceived effective task performance and was unlikely to be associated with ineffective effort (i.e. effort that is not perceived as achieving successful task performance) and fatigue (Hockey 2013).
Some studies have reported no correlation between different physiological measures of listening effort. For example, McMahon et al. (2016) reported no correlation between alpha power and pupil size when listening to noise-vocoded sentences with 6 (less intelligible) and 16 (more intelligible) channels in the presence of different levels of background noise. Increased cognitive resources were hypothesized to result in increased pupil size and EEG alpha power when listening to the 6-channel compared to the 16-channel vocoded sentences and as the SNRs became less favorable. Pupil size increased when listening to the less intelligible sentences (6 channels) at favorable signal-to-noise ratios (SNRs) only, while alpha power increased when listening to the more intelligible (16 channels) sentences at challenging SNRs only. The authors suggested that the lack of correlation might be due to different neurophysiological or attentional networks that modulate the activity of the physiological processes indexed by the different measures. Similarly, it was suggested that the often reported non-significant correlations between self-report and behavioural/physiological measures of listening effort may be because self-report and behavioural/physiological measures assess different aspects of listening effort/fatigue (Mackersie et al. 2015). However, whether self-report and behavioural/physiological measures do relate to different underlying aspects of listening effort/fatigue (and what these aspects are) has not been established.

**Inconsistent patterns between different groups of participants tested using the same measure**

Higher electroencephalography (EEG) alpha band (8-12 Hz) power was reported to index increased listening demands (e.g. Obleser et al. 2012; Dimitrijevic et al. 2017). However, patterns of change in alpha band power are inconsistent between studies. Petersen et al.
(2015) investigated EEG alpha power when: i) presenting speech in different levels of background noise and ii) manipulating the memory demands of a listening task by asking participants to memories 2, 4, or 6 digits. Alpha power increased in participants with normal-hearing or mild hearing loss but decreased in participants with moderate hearing loss in the more challenging listening conditions. Petersen et al. suggested that participants with moderate hearing loss had exerted maximal cognitive effort in the challenging listening conditions so that further increases in alpha power were not possible. The authors suggested that the decrease in alpha power is likely a result of participants “running out” of cognitive resources.

Ohlenforst et al. (2017b) examined peak pupil size when participants listened to sentences in the presence of two types of background maskers (single talker and stationary background noise) at a range of SNRs. A significant interaction was identified between participant group (normal hearing or hearing impairment) and both SNR and type of masker. For normal-hearing participants listening in both types of masker, increases in noise level led to increases in peak pupil size, up until the point where participants became unable to cope with the demands of the task and both pupil size and performance declined. Hence, peak pupil size was greatest at a narrow range of challenging SNRs. In comparison, hearing-impaired participants exhibited less pronounced changes in peak pupil size across SNR, and these changes were not significant. Additionally, in the presence of the single-talker masker, peak pupil size occurred at a more favorable SNR in the hearing-impaired participants than in those with normal hearing. The authors suggested the limited changes in peak pupil size in participants with hearing impairment might indicate that they had already recruited
significant cognitive resources at relatively favorable SNRs and so were less able to enlist further cognitive resources in the more challenging conditions. The different patterns of change in peak pupil size across both groups of participants suggest that listeners with normal hearing and listeners with hearing impairment allocate cognitive effort differently depending on task demands.

**Inconsistent findings between studies that used the same measures but different listening material**

Inconsistent findings have been reported for the same physiological measure of listening effort in different studies that recruited similar groups of participants but used different listening material; e.g. Obleser and Weisz (2012) and McMahon et al. (2016). Obleser and Weisz (2012) presented participants with words degraded using a noise vocoding technique. The authors reported decreased alpha power suppression (i.e. increased alpha power) when listening to speech with fewer acoustic details and suggested that it is an indication of increased mental activity which could provide insights to effortful listening. On the other hand, McMahon et al. (2016) reported higher alpha power when participants listened to more intelligible 12-channel noise-vocoded sentences compared to less intelligible 6-channel noise-vocoded sentences.

The different listening materials used across the studies might explain the contradictory findings. Obleser and Weisz presented participants with noise-vocoded single words while McMahon and colleagues presented participants with vocoded sentences. Kahneman’s model of attention (1973) suggests that listening is often an “automatic” process in ideal listening conditions. However, degradation of inputs limits the ability to map inputs to
automatic representations in the memory. Processing of sentences might be associated with increased limitations on the ability to automatically process speech inputs when trying to establish the relationship between the different items in the sentence. A non-linear relationship exists between task demands and listening effort (e.g. Ohlenforst et al. 2017b). Therefore, the difference in the listening demands associated with processing different speech materials complicates the ability to compare the results of different studies.

In summary, a variety of self-report, behavioural, and physiological measures of “listening effort” have been used in research studies. Although all measures have been interpreted in terms of “listening effort”, measures do not always agree well with each other, across participant groups, or between studies. The first explanation might be that measures are unreliable. Unreliable measures are unlikely to correlate with each other. The second explanation might be the inconsistencies in the listening tasks used across studies. Measures might correlate with each other if the same listening task was used. The third explanation might be that the various self-report, behavioural, and physiological measures may encompass different concepts that are related to listening effort, including arousal, attention, stress, and perceived difficulty (Pichora-Fuller et al. 2016). The various measures might also assess different processes or neural mechanisms involved in effortful listening such as the verbal working memory and attention-based performance monitoring (Peelle 2017). If there are multiple dimensions of “listening effort”, then multiple measures may be required for the assessment of listening effort. One final explanation might be that measures tap into underlying phenomena that are independent of the concept of listening effort. The use of the various measures of
listening effort was based on models and theories that provided links between increased listening demands and the potential measures. However, the absence of a gold standard for the assessment of listening effort limits the ability to confirm that the different measures relate to the concept of listening effort.

**Aims**

Multiple potential measures of listening effort were recorded simultaneously during a listening task that involved listening to digits in background noise in a large group of adult participants with a range of hearing levels. Measures included: (i) two self-report measures [NASA Task Load Index and the Visual Analogue Scale of Fatigue; VAS-F] (ii) one behavioural measure [reaction time] and (iii) three physiological measures [pupillometry, skin conductance, and EEG]. Other potential indicators of listening effort included performance on a speech in noise task and participants’ hearing level. The rationale for using each of the measures is provided below:

- Participants’ perception of listening difficulties should be the main interest in hearing rehabilitation. Therefore, the inclusion of a self-report measure in the design of this study was considered essential.
- Behavioural consequences of listening to degraded inputs include longer processing times (Gatehouse and Gordon 1990) and difficulty memorising the items presented (Rabbitt 1991). Behavioural measures can reveal increased cognitive demands before there is a decrement in performance.
- The use of EEG alpha power in the assessment of listening effort is based on the inhibition theory which suggests that increased alpha power is likely to occur in tasks requiring the retention of learned information or the suppression of
irrelevant inputs (Klimesch et al. 2007). Therefore, changes in alpha activity during a retention period where participants are required to memorise learned information was used as an index of listening effort. Increased alpha power while listening to speech in background noise was considered a potential indicator of effortful listening associated with the suppression of background noise. Alpha power in the baseline period was also considered a potential indicator of listening effort. According to Klimesch (2007), increased baseline alpha activity is an indicator of pre-task cortical engagement that predicts improved task performance. Including a predictor of task performance was motivated by recent reports suggesting that the accuracy of task performance can influence the experience of listening effort (Pichora-Fuller et al. 2016).

- Increased alertness results in increased pupil size (Kahneman 1973). Therefore, pupillometry has been traditionally considered an index of increased levels of alertness that might occur in demanding listening conditions (McGarrigle et al. 2014). On other occasions, increased pupil size has been considered an indication of increased task engagement associated with motivation and successful performance (e.g. Kuchinsky et al. 2014). Pupillometry provides an online method for momentary assessment of the changes in the ongoing neural activity during the performance of demanding tasks performance (Peelle 2017).

- Skin conductance provides an indication about the activity in the autonomic system. Activity in the sympathetic nervous system increases in demanding conditions in order to prepare the body to expend increased energy. This is referred to as the “fight or flight” response (McArdle et al. 2006). On this basis, skin conductance was considered a candidate measure of listening effort.
associated with listening to speech in demanding conditions (e.g. Mackersie and Cones 2011).

- Performance on a speech in noise task was considered a candidate measure of listening effort. Evidence suggests that performance on a speech task correlates with self-reported listening effort (e.g. Alhanbali et al. 2017b). The accuracy of performance on a listening task can influence listening effort, e.g. successful task performance can motivate further exertion of listening effort and vice versa. Therefore, performance accuracy might provide an indication about the experience of listening effort in individual participants (Ohlenforst et al. 2017b).

- Participants’ hearing level was also considered a candidate indicator of listening effort despite the lack of correlation with self-reported effort (Alhanbali et al. 2017a). As discussed above, the pattern of change in a number of listening effort measures (such as pupillometry and EEG alpha power) depends on participants’ hearing level (e.g. Petersen et al. 2015; Ohlenforst et al. 2017b).

The first aim of this study was to assess the reliability of the measures by testing a sub-group of participants on two separate occasions. A second aim was to assess the correlation between the different measures. The final aim was to use Factor Analysis (FA) to identify whether purported, reliable measures of listening effort assess similar or different underlying factor(s).

**Methods**

**Participants**
Participants were native English speakers recruited from the database of three UK National Health Service audiology departments and via flyers posted around the University of Manchester campus and through social groups. A total of 141 took part in the study. The data of 25 participants were not included in the factor analysis due to problems in the pupil or in the EEG data as will be described below. Therefore, data for 116 participants were included in the factor analysis. Participants’ age range was 55 to 85 years (M: 70, SD: 8), with 50.3% males. Hearing thresholds in the better ear of individual participants ranged from 10 to 77 dB HL over the frequencies 500, 1000, 2000, and 4000 Hz (M: 33, SD: 16.7). Participants with hearing level ≤ 30 dB HL at all frequencies (n: 37) were classified as having good hearing. The severity of hearing impairment for participants whose hearing level did not fall within the good category was classified according to a modified version of the British Society of Audiology classifications: mild (mean: 31-40 dB HL; n: 42, age: 68-83 years), moderate (mean: 41-70 dB HL; n: 29, age: 55-83 years), and severe (mean: 71-95 dB HL; n: 8, age: 61-83 years). Seventy participants were prescribed hearing aids by the NHS. All participants used behind-the-ear hearing aids with non-linear amplification fit according to the NAL-NL1 prescription target. Self-reported use was reported as “most of the day” for > 6 months. Participants performed the listening task with the hearing aid settings that they use in everyday life. The purpose of using everyday hearing aid settings was to measure listening effort in a cross-section of current hearing aid users, as was done by Alhanbali et al. (2017b). Therefore, we did not directly measure real ear gain to confirm audibility or if the hearing aids met the prescription target.
The sample size was determined on the basis of providing adequate statistical power to support a Factor Analysis (FA) i.e., a minimum of 5 to 10 participants per variable (Field 2009), with a minimum of one hundred participants in total (Floyd and Widaman 1995). The study was reviewed and approved by the National Research Ethics Services of South Central-Hampshire A, Research Ethics Committee reference: 15/SC/0113.

**Materials**

**Listening tasks**

The speech material was monosyllabic digits “1” to “9” from the Whispered Voice Test (McShefferty et al. 2013) recording of a male speaker. Bisyllabic number “7” was not included. The masker was unmodulated background noise. The noise started five seconds before the onset of the first digit and ended one second after the last digit had ended. Five seconds of noise is usually sufficient for the automatic noise reduction function in hearing aids to activate\(^3\). The SNR was determined using a sequence of 3 digits.

The listening task was performed in a sound-treated booth. The speech material was presented at a fixed level of 65 dB(A). Speech and background noise were both presented via loudspeakers at ±45° azimuth. Participants were seated facing a computer monitor. The height of the chair was adjusted to achieve the most comfortable setting for the participants with the head position supported using a chin rest.

The SNR required for each participant to identify 71% of the digits presented was established before performing the main listening task where listening effort was recorded.

---

\(^3\) Resound noise tracker II, White paper.
using the different measures. Refer to Alhanbali et al. (2017b) for details about establishing 71% criterion performance. In summary, the individualized SNR for each participant was established (for sequences of three digits) using a 2-down, 1-up, with a 2-dB step size adaptive procedure. This ensured equal intelligibility across participants and replicates approaches taken in previous studies (e.g. Mackersie et al. 2015; and Petersen et al. 2015). The mean SNR for criterion performance of 71% correct was -4 dB (SD: 5 dB).

Unlike the three-digit sequence used to determine individualized SNRs, the main study used sequences of six digits to maximize the cognitive demands of the task. Within each sequence of six digits, each digit was not repeated more than twice (e.g. 2 6 8 5 1 8). The listening task was a modified version of the Sternberg paradigm (Sternberg 1966) in which participants had to memorize speech material presented during a stimulus-free retention period based on similar paradigms described by Obleser et al. (2012) and Petersen et al. (2015). The listening task was programmed using SR research Experiment Builder software (SR Research version 1.10.1630, Mississauga, ON, Canada). Participants with hearing impairment performed the task with their hearing aids on.

Before the listening task, participants watched a documentary for 10 minutes (the baseline period) in order to acclimatize to the experimental setting and to obtain baseline values for skin conductance (see measures section below). The task started by presenting participants with the message “press ENTER when you are ready”. The word “Listen” then appeared on the screen and 5 seconds of unmodulated noise followed by the first sequence of six digits in noise were presented. A 3-second retention period followed, during which participants had to fixate on a cross while mentally rehearsing the digits. A digit then appeared on the screen and an audible pure tone was
presented to alert the participant to respond. Using a button box with “Yes” and “No” labels, participants responded with “Yes” if the digit on the screen was one of the digits they heard and with “No” if it was not. After responding, there was a recovery period of silence for 4 seconds before the start of a new trial to allow measures to return to baseline. The following instructions were verbally presented to each participant: “You are going to hear six numbers in background noise. After that, a cross will appear on the screen for three seconds of silence during which you are required to try to memorize the six numbers you heard. A single digit will then appear on the screen and a question mark beside it. You have to respond by pressing “yes” if the number on the screen was one of the six you heard and by pressing “no” if it was not. Try to respond as fast as you can. This is going to be repeated 50 times and the whole task will take around 15 minutes”. Before the start of the listening task, participants performed 10 practice trials of 6 digit sequences at their individualized SNR. The total number of experimental trials was 50. The overall duration of the listening task was around 15 minutes.

Figure 1 provides an outline of the sequence of events in each trial and the time periods used when analyzing the data obtained from the different measures, as will be discussed below.
Reliability of the measures

A subgroup of 30 participants performed a re-test one week after the first test session. According to Koo and Li (2016), a minimum of 30 samples (participants) are required to provide enough power for reliability testing. Both testing sessions were performed at the same time of day.

Listening effort and fatigue measures

Self-report scales

NASA Task Load Index and the VAS-F (Lee et al. 1991) were used for measuring self-reported listening effort and fatigue, respectively. NASA Task Load Index is a standardized measure for the assessment of perceived demands during task performance. The NASA Task Load Index consists of six items including: mental demand, physical demand, temporal demand, perceived performance, effort, and frustration. After performing the listening task, participants provided responses on a 20-step scale ranging from low
demand to high demand for each dimension. The score of each item was converted to a percentage. The total score was calculated based on the mean score of the items used.

The VAS-F consists of two subscales which are fatigue (12 items; e.g. fatigued, tired, and exhausted) and energy (6 items; e.g. active, energetic, and efficient). For each question, participants had to respond by choosing one number on a scale with two distinct points ranging from 0 to 10. For the fatigue items, larger numbers indicate more fatigue, while for the energy items larger numbers indicate more energy. The scales of the different items were converted so that they change in the same direction. The total score was calculated based on the mean score of the items used. Participants completed the VAS-F before and after performing the listening task. Final scores were based on the difference in mean VAS-F before and after performing the listening task. Although the duration of the listening task was only around 15 minutes, the development of fatigue was expected to occur due to the repetitive nature of the task that required participants to provide prompt responses (Hockey 2013).

**Behavioural measure**

Reaction time was used as the behavioural measure of listening effort. The time between the response prompt and participants’ response (button press) was recorded in milliseconds for both the correct and the incorrect responses and then averaged across all trials for each participant. Ideally, excluding reaction times of incorrect responses would increase the reliability of the measure. However, excluding incorrect responses involves removing trials to which incorrect responses were provided from the analysis of all of the other measures. This is a limitation for the analysis of EEG data where a
reasonable number of trials are required to obtain a good SNR. However, mean performance in the listening task was around 85% suggesting that excluding incorrect responses would not significantly affect the reaction time results. Reaction time information was exported through the SR research Experiment Builder software.

**Physiological measures**

**Pupillometry**

**Pupillometry recording**

Pupil sizes were measured using an Eyelink 1000 with a sampling rate of 1000 Hz. The eye tracker was connected to the same PC that was used to present the listening task. The desktop mount of the Eye link 1000 was used and the eye tracker was placed just below the lower edge of the computer monitor. Pupil size was measured based on the number of pixels in the pupil image captured by the camera which ranged from 100 to 1000 units with a precision of 1 unit corresponding to 0.01 to 5 mm pupil diameter. Pupil size was changed into mm by calculating the number of pixels in an artificial pupil with a known size.

The camera of the eye tracker was calibrated by asking participants to fixate on a black circle that periodically appeared at one of nine different coordinate positions on the computer monitor. Based on the luminance adjustment procedures reported in Zekveld et al. (2010), room lighting and screen brightness were adjusted for each participants to avoid floor/ceiling effects in pupil size. For each participant, pupil size was recorded in a bright (room brightness at 263 lux and screen brightness at 123 cd/m2), and a dark setting (room brightness at 0.28 lux and screen brightness at 0.0019 cd/m2). Room
lighting and screen brightness were then adjusted to achieve a pupil size that was in the middle range of the bright and the dark setting.

**Pupillometry data pre-processing**

In each trial, the pupil data included in the analysis ranged from the start of the speech stimulus and until the end of the 3-seconds retention period. Consequently, each epoch included the duration of the speech stimulus presentation plus the 3-seconds retention period (see Figure 1). The 3-second retention period was included in the analysis because of the lag of the peak pupil response that was observed in previous research (e.g. Piquado et al. 2010; Zekveld et al. 2010).

Pupil data were analyzed based on previous studies (Zekveld et al. 2010; Zekveld et al. 2011) using MATLAB (MathWorks Inc., version R2015a, MA, USA) scripts. Missing data points due to eye blinks were removed from the analysis. Based on Zekveld et al. (2011) and Ohlenforst et al. (2017b), trials with more than 15% of missing data points between the start of the baseline period to the end of the retention period were removed from the analysis. Linear interpolation using data points before and after the blink was applied to replace missing data points. Data were smoothed using 5-point moving average to remove any high-frequency artefacts. The mean number of trials lost for each participant was 5 (SD: 2). A total of 15 participants had more than 10 trials rejected due to problems such as drooping eyelids or diagnosed lazy eye and were thus excluded from the analysis.

**Pupillometry data analysis**
Once artefactual trials have been removed, the remaining trials were used to obtain two pupil outcome measures: i) peak pupil dilation amplitude, and ii) mean pupil dilation amplitude. Mean pupil size during the 1 second that preceded the presentation of the speech stimulus was used as a baseline (see Figure 1). Peak and mean pupil dilation were calculated relative to baseline i.e. in each trial, peak and mean pupil dilation were subtracted from mean pupil size during baseline. Mean and peak dilation were calculated for each trial. The final mean and peak pupil dilation for each participant was based on the average of the values obtained from all trials.

**EEG**

**EEG recording**

EEG was recorded using a Nexus-10 physiological recording system with the BioTrace software (Mind Media neuro and biofeedback system). EEG was sampled at 256 Hz with no online filtering. Increased alpha activity associated with increased listening effort has mainly been observed over the parietal lobe (Obleser and Weisz 2012; Obleser et al. 2012). Seven silver/silver chloride (Ag/AgCl) electrodes with a sintered surface were used. Three positive electrodes were therefore placed over parietal scalp regions to capture task-related alpha activity: Pz, P3, and P4 based on the international 10-20 system (Homan et al. 1987). The fourth positive electrode was placed at Cz. The positive electrodes placed at P3 and P4 were referenced to a negative electrode placed at the left ear lobe. The positive electrodes placed at Pz and Cz were referenced to a negative electrode placed on the right ear lobe. The ground electrode was placed at the forehead. Before placing the EEG electrodes using conductive paste, the skin was prepared using an abrasive gel. Electrode impedance was kept below 5 ohm.
**EEG data pre-processing**

EEG data were processed using EEGLAB tool box (Delorme and Makeig 2004). The first 0.5 seconds of any pre-determined time periods (noise/speech/retention) were excluded from the analysis so as to avoid any stimulus onset or offset activity (Petersen et al. 2015). Epoched data were filtered between 5 and 45 Hz using EEGLAB (Petersen et al. 2015). The filter function in EEGLAB applies padding to the epoched data which controls for edge effects or artefacts that might occur as a result of filtering epoched data. Trials containing artefacts, including blinks, saccadic eye movements, or EMG activity, were removed from further analysis. Participants’ data with more than 20% rejected trials were not included in the analysis (Cohen 2014). The mean number of trials lost for each participant was 7 (SD: 3). A total of 10 participants had more than 10 trials contaminated with artefacts and were excluded from further analysis.

**EEG time-frequency analyses**

Time-frequency decomposition using Morlet wavelet convolution was applied to the data. Complex wavelet convolution was performed to quantify changes in event-related band power (ERBP; Nourski et al. 2009) over the time periods outlined in Figure 1 (-0.7 to 13 seconds around the onset of a trial). ERBP for the retention period (top panel of Figure 2) was estimated for each center frequency from 5 to 20 Hz in 1 Hz steps. Power estimates during the retention period were calculated relative to power estimates during the pre-stimulus baseline period (-0.6 to -0.1 seconds of the stimulus onset) (Petersen et al. 2015). Power estimates were also calculated during the speech presentation period (bottom panel of Figure 2) but used a different baseline defined during the presentation
of noise alone to ensure that any increase in alpha activity is in response to the presentation of speech and not merely a response to noise (-0.6 s to -0.1 seconds before the speech onset) (Dimitrijevic et al. 2017). Power estimates during the pre-stimulus baseline period (-0.6 to -0.1 seconds of the stimulus onset) were calculated and included in the FA to determine whether pre-stimulus alpha predicted task performance (Klimesch 2007).

The alpha ERBP was quantified for each individual participant in the center frequencies ranging from 8-13 Hz using EEGLAB tool box. To do so, trial data were convoluted with a family of 3 Morlet waves (default setting of EEGLAB). Alpha power was calculated during the pre-stimulus baseline (-0.6 to -0.1 seconds) and the retention period (9.5 to 12 seconds into the trial). Alpha power was also calculated during the noise baseline period (-0.6 to -0.1 seconds before the speech onset), and during the presentation of the speech (5.5 to 8.5 seconds into the trial). For each center frequency and each time point, power estimates were obtained by calculating the logarithm of the mean power during the retention period over the mean power during the baseline period. Alpha power was then averaged across the frequencies 8 to 13 Hz. Alpha power was calculated in each trial and then averaged across trials for each participant.

To visualize a time-frequency representation of the data (Figure 2), customized MATLAB scripts developed by Nourski et al. (2009) were used. Time-frequency decomposition using Morlet wavelet convolution \(2\pi f_0 \sigma = 7\) (Petersen et al. 2015) was applied to the data averaged across all participants. The entire filtered frequency range, i.e. 5 to 45 Hz is
not presented in Figure 2 to allow a better visualization of changes in alpha power (8 to 13 Hz).

![Figure 2](image)

**Figure 2.** Mean change in alpha power across participants and trials. The temperature scale represents changes in event-related band power in decibels (dB). The top panel shows changes in alpha activity during the retention period relative to baseline alpha activity in the recovery period (Petersen et al., 2015) i.e. before the noise is presented. The bottom panel shows changes in alpha activity during the speech presentation period relative to alpha activity during the last second of unmodulated noise i.e. the period of noise alone that preceded the presentation of the first spoken digit. Dashed boxes represent the time periods included in the analysis. (n =116).

**Skin conductance**

Recordings of skin conductance and EEG were performed simultaneously via separate channels in the Nexus-10. Skin conductance was sampled at 32 Hz. Two silver/silver chloride (Ag/AgCl) electrodes were attached to the index and the middle finger of the
participant’s non-dominant hand. Participants were instructed to keep their hand facing palm-up to minimize artefacts resulting from hand movement or any pressure applied on the electrodes.

Skin conductance data were extracted through the Biotrace software. The epoch of each trial commenced from the start of the stimulus and terminated at the end of the retention period. We did not include the 4-second recovery period in the skin conductance analysis as participants did not do any mental task during that period.

In order to account for the individual differences in baseline skin conductance, mean skin conductance for each participant across all trials was corrected to baseline. Pilot testing indicated that it took around 3 minutes for the skin conductance values to settle. As a result, average skin conductance value in the 7 minutes that preceded task performance (while watching the documentary) was used as a baseline. Mean skin conductance across trials was subtracted from mean skin conductance in the baseline period. The value resulting from the subtraction was then divided by mean skin conductance in the baseline period.

**Statistical analysis**

The data were not normally distributed and were therefore summarized using median and inter-quartile ranges (IQR), and analysis involved nonparametric tests.

Test re-test reliability was assessed using Spearman’s correlation coefficient (consistency of the results across the testing sessions) and Inter-class Correlation Coefficient (ICC; test re-test reliability). ICC estimates and 95% confidence interval were calculated based on an absolute agreement one way random effects; ICC1 based on Shrout and Fleiss (1979).
ICC1 is sensitive to differences in means between the observations and is a measure of absolute agreement. Each session for each participant can be considered a separate condition due to differences in aspects such as electrode placement or how alert the participant is on the day of testing. Therefore, every session can be regarded as being conducted by a separate "rater" or "judge" suggesting that ICC1 is likely the most appropriate to use for these data (Shrout and Fleiss 1979). The correlations between the different variables were investigated using Spearman’s correlation coefficient. The correlation between each of the different variables and age was also investigated.

The suitability of the data for a FA was investigated using Kaiser–Meyer–Olkin measure of sampling adequacy (KMO) test and Bartlett’s test of sphericity (Field 2009). FA included only the measures that were shown to have good re-test reliability (see later). Factors were identified based on eigenvalues greater than one (Field 2009). Oblique rotation was applied to the data to identify how the measures load into distinct factors (Field 2009). Multiple parameters of EEG and pupillometry were included in the Factor Analysis since these might tap into independent aspect of increased listening effort. For example, increased alpha activity during the retention period was considered an indication of increased demands on the working memory (e.g. Petersen et al. 2015) whereas increased alpha activity during the speech presentation period was considered an indication of suppression of background noise (McMahon et al. 2016). Furthermore, measures of EEG alpha during the baseline period may be predictive of task performance (Klimesch 2007).

Results

Test-retest reliability
Figure 3 shows the relationship between the test and re-test results. Spearman’s correlation coefficients and ICC with 95% CI for the different measures are summarized in Table 2. Spearman’s correlation coefficients indicated excellent consistency across the testing sessions for all measures except for skin conductance, which was moderately consistent, and VAS-F which had poor consistency. Pupillometry had good to excellent reliability, EEG (alpha power) had moderate to excellent reliability, reaction time had moderate to good reliability, skin conductance had poor to good reliability, NASA Task Load Index had moderate to excellent reliability, and VAS-F had poor to moderate reliability based on the ICC classification suggested by Koo and Li (2016). Skin conductance and VAS-F were not included in the Factor Analysis due to poor test re-test reliability.

Table 2. Correlation coefficients between the test and the retest sessions of the different measures and results of ICC calculation with confidence intervals.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Spearman’s correlation coefficient (p value)</th>
<th>Intraclass Correlation Coefficient</th>
<th>95% Confidence Interval</th>
<th>Classification based on ICC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean pupil size</td>
<td>rs = 0.80 (p &lt; 0.05)</td>
<td>0.85</td>
<td>0.71 - 0.93</td>
<td>Moderate to excellent</td>
</tr>
<tr>
<td>Peak pupil size</td>
<td>rs = 0.84 (p &lt; 0.05)</td>
<td>0.90</td>
<td>0.80 - 0.95</td>
<td>Good to excellent</td>
</tr>
<tr>
<td>Reaction time</td>
<td>rs = 0.77 (p &lt; 0.05)</td>
<td>0.74</td>
<td>0.54 - 0.86</td>
<td>Moderate to good</td>
</tr>
<tr>
<td>Skin conductance</td>
<td>rs = 0.55 (p &lt; 0.05)</td>
<td>0.57</td>
<td>0.27 - 0.78</td>
<td>Poor to good</td>
</tr>
<tr>
<td>EEG</td>
<td>rs = 0.80 (p &lt; 0.05)</td>
<td>0.81</td>
<td>0.65 - 0.90</td>
<td>Moderate to excellent</td>
</tr>
<tr>
<td>NASA</td>
<td>rs = 0.80 (p &lt; 0.05)</td>
<td>0.81</td>
<td>0.64 - 0.90</td>
<td>Moderate to excellent</td>
</tr>
<tr>
<td>VAS-F</td>
<td>rs = 0.25 (p &gt; 0.05)</td>
<td>0.27</td>
<td>-0.10 - 0.57</td>
<td>Poor to moderate</td>
</tr>
</tbody>
</table>
Figure 3. Correlations between the test (x-axis) and re-test (y-axis) data (n = 30). RT: reaction time, SC: skin conductance.

**Descriptive statistics**

Mean performance on the listening task (that required participants to memorize the 6 digit sequence) was 93% (SD: 4) for participants with good hearing, 89% (SD: 7) for
participants with mild hearing loss, 87% (SD: 6) for participants with moderate hearing loss, and 82% (SD: 6) for participants with severe hearing loss. The median score and inter-quartile range (IQR) for the NASA Task Load Index were 34.16% (IQR: 26.25). For VAS-F, the values were 6.50% (IQR: 17.96). For reaction time the values were 1945.86 milliseconds (IQR: 540.71) and for skin conductance, 0.25 µS (IQR: 0.30).

**Pupillometry**

Visual inspection of pupillary changes during the recovery period suggested that pupil size returned to baseline before the start of a new trial. Figure 4 shows mean change in pupil size across all participants (n: 116) and trials. Pupil size increased significantly relative to baseline as participants attended to the speech, and reached a peak towards the end of the 4-second speech stimulus. Median pupil size across participants was 0.16 mm, IQR = 0.54. Median peak pupil size was 1.18 mm, IQR = 0.92.

![Figure 4](#)
**EEG**

Figure 2 shows the mean ERBP (Nourski et al. 2009) across participants (n:116) and trials. The top panel represents mean ERBP during the retention period relative to baseline during the recovery period. The bottom panel represents mean ERBP during the presentation of the digits in noise relative to baseline during the presentation of the noise only. Changes in ERBP are represented by the temperature scale which ranges from -5 to 5 dB. Figure 2 suggests an increase in alpha activity towards the end of the retention period and an increase in alpha activity during speech presentation (8-13 Hz; highlighted by black dashed box). A Wilcoxon rank test was used to establish whether alpha power during the retention period and during the speech presentation period significantly increased compared to their respective baselines. Increased alpha activity was identified during speech presentation period only (0.5-4 seconds); \( z = -2.30, p < .05 \). Median alpha power during speech presentation across participants was 0.17 dB, IQR = 1.99. Median alpha power during retention was -0.97 dB, IQR = 1.90. Figure 2 also shows that alpha activity during the baseline period was around 0 dB suggesting that the recovery period was long enough for alpha activity to return to baseline.

**Correlations and Factor Analysis**

Some weak correlations were identified between the measures (Table 3). Age was weakly correlated with SNR \( (r = 0.29, p < 0.05) \) and not correlated with any other measure. Therefore, age was not included in the Factor Analysis. FA involved 10 variables:

1. NASA Task Load Index,
2. SNR,
3. reaction time,
4. mean pupil size,
5. peak pupil size,
6. EEG alpha during baseline period,
7. EEG alpha during retention period,
8. EEG alpha during speech presentation,
9. hearing level, and
10. performance accuracy

Results of a KMO test (0.59) indicated the adequacy of the sample size for a FA (Field 2009). According to Field (2009), KMO values below 0.50 are unacceptable for a Factor Analysis. Bartlett’s test of sphericity X²(45) = 158.214, p < 0.001, indicated that correlations between the variables were sufficient for a FA. FA yielded 4 factors with eigenvalues > 1 that explained about 67% of the total variance (Table 4).
Table 3. Correlation between the measures.

<table>
<thead>
<tr>
<th>VAS-F</th>
<th>NASA</th>
<th>VAS-F</th>
<th>Reaction time</th>
<th>Skin conductance</th>
<th>Performance</th>
<th>PTA</th>
<th>Alpha baseline</th>
<th>Alpha retention</th>
<th>Alpha speech</th>
<th>Mean pupil</th>
<th>Peak pupil</th>
<th>SNR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spearman’s r</td>
<td></td>
<td>.313</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td></td>
<td>.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reaction time</td>
<td>Spearman’s r</td>
<td></td>
<td>.009</td>
<td>.074</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td></td>
<td>.925</td>
<td>.430</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skin conductance</td>
<td>Spearman’s r</td>
<td></td>
<td>.224</td>
<td>-.093</td>
<td>-.137</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td></td>
<td>.016</td>
<td>.320</td>
<td>.143</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance</td>
<td>Spearman’s r</td>
<td></td>
<td>-.122</td>
<td>-.112</td>
<td>-.090</td>
<td>-.147</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td></td>
<td>.228</td>
<td>.272</td>
<td>.374</td>
<td>.147</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTA</td>
<td>Spearman’s r</td>
<td></td>
<td>.161</td>
<td>.051</td>
<td>.186</td>
<td>.157</td>
<td>-.598**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td></td>
<td>.095</td>
<td>.602</td>
<td>.053</td>
<td>.103</td>
<td>.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alpha baseline</td>
<td>Spearman’s r</td>
<td></td>
<td>-.212*</td>
<td>-.162</td>
<td>.105</td>
<td>-.268*</td>
<td>.044</td>
<td>-.109</td>
<td>-.267**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td></td>
<td>.023</td>
<td>.084</td>
<td>.266</td>
<td>.004</td>
<td>.664</td>
<td>.259</td>
<td>.004</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alpha retention</td>
<td>Spearman’s r</td>
<td></td>
<td>-.132</td>
<td>-.037</td>
<td>.034</td>
<td>-.252*</td>
<td>.111</td>
<td>-.191</td>
<td>.060</td>
<td>.288**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td></td>
<td>.166</td>
<td>.703</td>
<td>.721</td>
<td>.008</td>
<td>.286</td>
<td>.051</td>
<td>.533</td>
<td>.002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alpha speech</td>
<td>Spearman’s r</td>
<td></td>
<td>.026</td>
<td>-.258*</td>
<td>-.072</td>
<td>.243*</td>
<td>-.057</td>
<td>.068</td>
<td>-.009</td>
<td>-.082</td>
<td>-.123</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td></td>
<td>.782</td>
<td>.095</td>
<td>.446</td>
<td>.009</td>
<td>.573</td>
<td>.480</td>
<td>.927</td>
<td>.383</td>
<td>.199</td>
<td></td>
</tr>
<tr>
<td>Mean pupil</td>
<td>Spearman’s r</td>
<td></td>
<td>.093</td>
<td>-.046</td>
<td>-.096</td>
<td>.194</td>
<td>.137</td>
<td>-.085</td>
<td>.186</td>
<td>-.081</td>
<td>-.071</td>
<td>.592**</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td></td>
<td>.323</td>
<td>.626</td>
<td>.308</td>
<td>.038</td>
<td>.177</td>
<td>.382</td>
<td>.046</td>
<td>.389</td>
<td>.462</td>
<td>.000</td>
</tr>
<tr>
<td>Peak pupil</td>
<td>Spearman’s r</td>
<td></td>
<td>.019</td>
<td>.051</td>
<td>.183</td>
<td>.064</td>
<td>-.466**</td>
<td>.711**</td>
<td>-.174</td>
<td>.040</td>
<td>-.003</td>
<td>-.158</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td></td>
<td>.838</td>
<td>.591</td>
<td>.053</td>
<td>.501</td>
<td>.000</td>
<td>.000</td>
<td>.065</td>
<td>.675</td>
<td>.972</td>
<td>.995</td>
</tr>
<tr>
<td>SNR</td>
<td>Spearman’s r</td>
<td></td>
<td>.019</td>
<td>.051</td>
<td>.183</td>
<td>.064</td>
<td>-.466**</td>
<td>.711**</td>
<td>-.174</td>
<td>.040</td>
<td>-.003</td>
<td>-.158</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td></td>
<td>.838</td>
<td>.591</td>
<td>.053</td>
<td>.501</td>
<td>.000</td>
<td>.000</td>
<td>.065</td>
<td>.675</td>
<td>.972</td>
<td>.995</td>
</tr>
<tr>
<td>Age</td>
<td>Spearman’s r</td>
<td></td>
<td>-.038</td>
<td>-.068</td>
<td>.170</td>
<td>.015</td>
<td>-.005</td>
<td>.154</td>
<td>.023</td>
<td>-.041</td>
<td>.102</td>
<td>.083</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed)</td>
<td></td>
<td>.705</td>
<td>.497</td>
<td>.089</td>
<td>.879</td>
<td>.961</td>
<td>.131</td>
<td>.822</td>
<td>.684</td>
<td>.320</td>
<td>.408</td>
</tr>
</tbody>
</table>
Table 4. Percentage of the variance explained by each factor.

<table>
<thead>
<tr>
<th>Component</th>
<th>Extraction Sums of Squared Loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
</tr>
<tr>
<td>1</td>
<td>2.425</td>
</tr>
<tr>
<td>2</td>
<td>1.727</td>
</tr>
<tr>
<td>3</td>
<td>1.488</td>
</tr>
<tr>
<td>4</td>
<td>1.026</td>
</tr>
</tbody>
</table>

Table 5 provides details of the loading of each variable onto the different factors. To facilitate the interpretation of the data, low loadings of less than 0.30 are not shown (Field 2009). Both structure and pattern matrices yielded similar results, so only the pattern matrix is reported here. Oblique rotation resulted in performance accuracy, hearing level, baseline alpha power during the pre-stimulus period, and SNR loading into Factor 1. Mean pupil size and peak pupil size loaded into Factor 2. Alpha power (at baseline and during stimulus presentation and the retention period), and self-reported listening effort into Factor 3. Reaction time, self-reported effort, and baseline alpha power loaded into Factor 4. Table 6 is the correlation matrix between the four factors identified. The weak correlations suggest that the factors identified are independent, i.e. orthogonal.
Table 5. Pattern Matrix: factor loadings of the variables to each factor.

<table>
<thead>
<tr>
<th>Component</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reaction time</td>
<td></td>
<td></td>
<td>.838</td>
<td></td>
</tr>
<tr>
<td>NASA</td>
<td></td>
<td>-.365</td>
<td></td>
<td>.507</td>
</tr>
<tr>
<td>Performance</td>
<td>-.806</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTA</td>
<td>.851</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline Alpha</td>
<td></td>
<td>-.367</td>
<td>-.468</td>
<td>-.334</td>
</tr>
<tr>
<td>power</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alpha power</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>during retention</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alpha power</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>during speech</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean pupil size</td>
<td></td>
<td></td>
<td>.917</td>
<td></td>
</tr>
<tr>
<td>Peak pupil size</td>
<td></td>
<td></td>
<td>.900</td>
<td></td>
</tr>
<tr>
<td>SNR</td>
<td>.849</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6. Correlation matrix between the factors.

<table>
<thead>
<tr>
<th>Factor</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-.007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.017</td>
<td>-.060</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.227</td>
<td>.018</td>
<td>-.042</td>
</tr>
</tbody>
</table>

Discussion

Test-retest reliability

All measures except skin conductance and VAS-F had good-to-excellent reliability. Although results of a Wilcoxon rank test suggested the absence of a significant difference in mean skin conductance across both test sessions, it was only moderately reliable. Skin conductance is more sensitive to emotional factors than the other measures used in this study (Hogervorst et al. 2014) and this might explain the higher variability for this measure. Results of pilot testing indicated that skin conductance has low time resolution; i.e. it changes slowly over time (3 minutes were required for the measure to stabilize during pilot testing). Low time resolution might have also contributed to the low reliability of the measure. VAS-F also showed poor test-retest reliability. VAS-F scores were based on the difference (after − before) in participants’ rating of fatigue. On
inspection of the individual measures before and after task performance, pre-task fatigue ratings were unreliable (ICC: 0.26), while post-task fatigue ratings showed good reliability (ICC: 0.83). Poor reliability of the baseline fatigue measures could have been influenced by multiple factors that are difficult to control across participants: factors associated with various aspects of daily life prior to commencement of the test sessions. In contrast, fatigue rating after the task was most likely based on task performance and thus was more likely to be reliable. This suggests that a direct measure of change in fatigue (i.e. having participants rate fatigue after performing a listening task) may be a better measure than deriving the difference between two states (i.e. comparing self-reported fatigue before and after task performance). This is consistent with Gatehouse (1999) who demonstrated the improved discriminatory capability of using a single change measure. To our knowledge, the present study is the first to report the reliability of all of these measures of listening effort.

**Correlation between the measures**

Despite good reliability of the measures, they were only weakly correlated with each other. The use of different measures was motivated by theories and models which suggested their sensitivity to increased listening demands. Therefore, weak correlations between reliable measures suggest that they tap into different underlying dimensions that might relate to the concept of listening effort which supports the notion that listening effort is a multidimensional process involving multiple cortical regions (Peelle 2017). Processing degraded speech during effortful listening involves multiple cortical regions responsible for cognitive processes essential for understanding degraded inputs such as verbal working memory and attention-based performance monitoring (Peelle
In fact, processing degraded speech activates an even greater array of cortical regions (such as those involved in error monitoring) than are responsible for processing of clear speech (Peelle 2017). It is important to bear in mind that one or more of the factors identified might not be related to listening effort. Some of the factors might be indicators of processes associated with listening to degraded speech without being necessarily related to the concept of listening effort. For example, increased pupil size might be an indication of increased task engagement rather than effortful listening as will be discussed below.

Multi-dimensional cognitive psychological models of attention might provide a relevant framework for understanding the multidimensionality of listening effort. Kahneman’s model of attention (1973) suggests that effort is the consequence of mismatch between the cognitive demands of a task and the supply of cognitive resources. Pichora-Fuller (2016) suggested that Kahneman’s model of attention is relevant to effortful processing of degraded speech because it suggests that the presence of a perceptual deficit is a direct cause of increased effort. Kahneman’s model suggests that multiple dimensions control the allocation of cognitive resources for task performance such as the levels of arousal and the evaluation of task demands with respect to capacity. Therefore, it was possible to hypothesize that multiple factors influence the experience of effort. In an elaboration on Kahneman’s model of attention, Pichora-Fuller (2016) suggested that measures of listening effort might tap to the multi-dimensional attention-related outputs of Kahneman’s model, (including cognitive-behavioural changes, changes in brain activity, changes in autonomic nervous system activity, and in self-report measures) in their Framework for Understanding Effortful Listening (FUEL). The multi-dimensional attention-
related outputs described in the FUEL are consistent with the finding that different measures of listening effort tap into different underlying dimensions.

Identifying how different measures of listening effort group into independent factors will aid our understanding of the underlying dimensions of listening effort. Factor Analysis suggested that the 10 variables in this study related to four underlying factors, as follows.

**Factor One: Performance**

Hearing level, SNR, performance accuracy, and pre-stimulus baseline alpha power loaded into one factor: better hearing is associated with decreased SNR required for 71% criterion performance, better task performance, and greater resting (baseline) alpha power. Alpha power has similar (relatively weak) loading to Factors One, Three, and Four. This suggests alpha power during the baseline period has an influence of the different factors but does not explain much variability in each of them.

Our interpretation of Factor One is that it relates primarily to task performance. The presence of hearing impairment limited participants’ ability to identify the digits. Therefore, participants with poorer hearing required more favorable SNR to achieve 71% criterion performance. It is plausible that hearing level is related to task performance; i.e. participants’ ability to recall correct digit from sequence of 6 presented in noise. Intelligibility was equalized across participants. However, this does not imply equal effort. Therefore, hearing-impairment adversely impacted the ability to memorize the digits, even though the spoken stimuli were intelligible (Rabbitt 1991).
Greater baseline alpha activity is thought to be an indicator of pre-task cortical engagement and a predictor of the ability to perform a task (Klimesch et al. 2007). This is supported by the findings of studies that reported a correlation between better performance of IQ tests and higher amplitude of alpha power during a baseline period (Doppelmayr et al. 2002). Lower resting alpha activity may also reflect age-associated declines in cognition (Klimesch et al. 1999). Lower cognitive capacity (indexed by lower baseline alpha power) associated with hearing impairment may be linked to poorer task performance (Baltes and Lindenberger 1997).

**Factor Two: Task engagement**

Peak and mean pupil size loaded into a single independent factor which we interpret as being related to “task engagement”. The finding that pupillometry loaded into an independent dimension was not surprising since previous research findings suggested that: (i) pupillometry and self-report measures assess independent underlying dimensions (e.g. Wendt et al. 2016; Zekveld et al. (2011), (ii) changes in pupil size do not correlate with hearing level (Zekveld et al. 2011), and (iii) independent attentional networks control pupillometric and EEG activity (McMahon et al. 2016). Increased pupil size is an indication of increased arousal and alertness that might occur as a result of increased task demands (Kahneman and Beatty 1966). However, increased pupil size with increased listening demands might not always be an indication of a negative experience or an indication of “ineffective” effort. It has been reported that increased pupil size is an indication of increased attentiveness and task engagement (e.g. Kuchinsky et al. 2014; Wendt et al. 2016; Ohlenforst et al. 2017b).
**Factor Three: Cognitive processing**

EEG alpha power during speech presentation, retention of stimulus information, and baseline, loaded with self-reported effort into one factor, which we call “cognitive processing”. Self-reported listening effort loaded into Factor Three and Factor Four, with stronger factor loading to Factor Four (factor loading to Factor Three: 0.37, factor loading to Factor Four: 0.51). Therefore, self-reported effort can be considered more relevant to Factor Four. Alpha power is considered an indicator of cognitive processes and has been observed to increase with more demanding tasks, e.g. listening to degraded speech and increased memory demands (Obleser et al. 2012). Consistent with some previous work, here, we found increased alpha power during auditory stimulus presentation (McMahon et al. 2016) and to a lesser extent, during the retention period (Obleser et al. 2012). We could not replicate the relation between hearing level and changes in alpha power during the retention period reported by Petersen et al. (2015), possibly because their task tracked 80% correct performance and we adjusted intelligibility to a criterion performance of 71%, which would have been more challenging for our participants. Listening to speech in background noise is also expected to be associated with increased self-reported listening effort. However, in the present study, alpha power during both auditory stimulus presentation and stimulus retention loaded in the opposite direction to self-reported effort: greater EEG alpha power was associated with lower self-reported effort.

Participants’ perceived performance in the listening task might affect their experience of listening effort (Pichora-Fuller et al. 2016; Ohlenforst et al. 2017b). In this study, the differences in baseline alpha power were found to be related to participants’
performance, even after controlling for the effects of hearing level by carrying out a partial correlation analysis that controlled for the effect of hearing level (the correlation between baseline alpha power and performance remained significant, \( p=0.005 \)). Results also suggest that participants with increased alpha power in the baseline had increased alpha power during both the speech presentation and the retention periods. One interpretation of these effects may be that participants who perceived that they were performing better in the listening task were also more motivated to exert increased listening effort (Pichora-Fuller et al. 2016), as reflected by the increased alpha power during both the speech presentation and retention periods. In addition, perceived success in task performance might have maintained motivation and the task perceived as rewarding and relatively less effortful.

The effects of perceived performance may account for the relationship between greater EEG alpha power (during speech presentation and retention) and lower self-reported effort. More research is required to gain a better understanding on how hearing impairment, adverse listening conditions, and participants’ performance influence changes in alpha power. Changes in alpha power during a listening task might be an objective method that provides insights into individuals’ perception of listening difficulty.

**Factor Four: Behavioural consequences**

Reaction time and self-reported effort loaded into Factor Four in the same direction. Alpha power during the baseline period loaded into Factor Four in an opposite direction to the other variables Participants with increased baseline alpha power (indication of improved ability to perform the task) experienced less cognitive demands as suggested by
decreased reaction time and decreased self-reported listening effort. Self-reported effort loaded more strongly into Factor Four than Factor Three suggesting that perceived effort is more influenced by the behavioural outcomes of effortful listening. It has been previously reported that increased listening demands result in slower reaction times (Houben et al. 2013). Poorer hearing may increase listening demands even when intelligibility is accounted for (Rabbitt 1991), and so lead to slower reaction times. The design of the listening task in the current study was based on the protocol of previous studies that measured reaction time after a stimulus-free retention period (e.g. Obleser et al. 2012; Petersen et al. 2015). However, the interpretation of reaction time in the context of the present study is not straightforward because the design of our study did not require participants to provide responses immediately after stimulus presentation. In the current paradigm, the time to respond to the memory prompt was timed. Therefore, reaction time could be related to the memory and not the listening task per se. The assumption that reaction time can be considered an index of listening effort is also limited by the fact that it is not a process-pure measure; i.e. multiple aspects can influence the speed of processing such as individuals’ age (Pichora-Fuller et al. 2016). The limitations associated with the use of reaction time in the present work might have contributed to the finding that Factor Four explained the least variability in the data (Table 4).

The fact the self-reported effort loaded into the same dimension might provide confidence that reaction time is related to listening effort. However, it is possible that participants’ responses to the self-report measure were influenced by their perceived performance (i.e. how fast they were able to provide responses) and this might have
contributed to both variables loading into one factor. Again, the relationship between performance and baseline alpha power might be the reason for baseline alpha power loading into Factor Four. The opposite loading direction suggests that participants with increased baseline alpha power (an indication of pre-task cortical engagement) were more able to provide quicker responses to the task and experienced less perceived listening effort.

**Limitations**

Equalizing the intelligibility at a criterion performance of 71% (in an attempt to replicate previous research) might have reduced differences in listening effort between participants with different levels of hearing impairment. However, during pilot testing, it was not possible to identify a single SNR where all participants were able to achieve a similar level of performance. The SNR at which the participants performed the listening task might have also been unrealistically challenging compared with real-life situations. The SNRs at which participants were tested actually represent listening situations that individuals would usually avoid (Smeds et al. 2015). The purpose of choosing challenging SNRs in lab settings is usually to avoid ceiling effects. However, this might not be essential if measures of listening effort are sensitive to changes in performance at SNRs that are more representative of real life situations. For example, Krueger and colleagues (2017) reported that a measure of self-reported listening effort indicated that the use of hearing aids or noise reduction algorithm results in decreased listening effort when participants listened to speech in favorable listening conditions. Future work should consider obtaining simultaneous recordings while participants perform a listening task at noise levels that are representative of real-life situations. Evidence suggests that performance
accuracy can influence the results obtained using different listening effort measures. For example, performing a listening task in challenging listening conditions might limit task engagement and demotivate individuals to exert further effort. Therefore, performing the listening task at more favorable SNRs would also allow an in-depth investigation of the effect of task performance on listening effort. The use of multi-talker background noise might have been more representative of real-life situations compared to unmodulated noise. Greater listening effort has been reported in the presence of a single-talker “informational” masker compared to unmodulated noise (Koelewijn et al. 2012). Future work should consider the aforementioned factors when obtaining simultaneous recordings of the different measures. Consideration of these methodological factors might eliminate any effect they might have had on the factors identified, e.g. participants’ performance in the listening task was considered to have influenced the loading direction of a number of the measures.

Asking participants to memorize spoken digits may not have been sufficiently demanding to be sensitive to differences in effort. A task requiring manipulation of speech information might tap into aspects of everyday listening that might be more cognitively demanding and more sensitive to changes in listening effort. For instance, Rabbitt (1991), did not identify an effect of hearing impairment on participants’ ability to correctly repeat words; however, hearing impairment reduced the participants’ ability to correctly recall the words.

The relatively low KMO value suggests that the interpretation of the FA results should be treated with caution and justifies the need for independent replication of our current
study. Finally, the improved discriminatory capability of using a direct measure of change (e.g., in self-reported fatigue) is an important psychometric principle to be borne in mind when designing self-report measures and evaluating change.

Conclusions

Here, for the first time, we directly compared simultaneously recorded multi-modal measures of listening effort. The measures used in this study tap into independent dimensions that we interpreted as: i) performance, i.e. the ability to correctly respond to a task requiring detecting and processing auditory information, ii) task engagement, i.e. being aroused to detect the occurrence of an auditory stimulus, iii) processing i.e. cognitive processing of auditory information, and iv) behavioural consequences. This doesn’t mean, however, that these four factors are directly related to listening effort. For example, pupillometry can be used to measure task engagement and this is not a direct measure of effort. This suggests that the different measures should not be used interchangeably. The present work suggests the importance of careful and in-depth consideration of the aspect of increased listening demands of interest before choosing an appropriate measure. As such, the findings of the present study have widespread implications for both research and clinical practice.

Acknowledgements

The study was supported by the NIHR Manchester Biomedical Research Centre. We thank Dr Andrew Dimitrijevic for assistance in developing the research protocol. We also thanks Dr Mohammad Sobuh, Dr Emanuele Perugia, and Dr Garreth Prendergast for their assistance in the data analysis.
References


CHAPTER SIX

GENERAL DISCUSSION
CHAPTER SIX
GENERAL DISCUSSION

This PhD thesis aimed to answer the following research questions:

1-Do individuals with hearing impairment report increased listening effort and fatigue compared to controls with good hearing? (Study One)

2- Does hearing handicap, speech recognition in noise, and hearing level correlate with self-reported listening effort and fatigue? (Study Two)

3- Are potential measures of listening effort reliable? (Study Three)

4- Do potential measures of listening effort correlate with each other? (Study Three)

5- Do potential measures of listening effort tap into the same underlying concept? (Study Three)

The purpose of this chapter is to provide a general discussion after completing all the studies and not to repeat the discussion already provided in the individual manuscripts.

Results of Study One confirmed the hypothesis that individuals with hearing impairment report increased levels of listening effort and fatigue compared to controls with good hearing. Our findings justify the importance of considering aspects of listening effort and fatigue in the hearing rehabilitation process. With the exception of the three effort-related questions in the SSQ Hearing Scale (Gatehouse and Noble 2004), scales that assess self-reported listening effort and fatigue in the daily life of individuals with hearing impairment do not exist. It is possible to use generic fatigue scales as have been previously done in a number of research studies, e.g. Hornsby and Kipp (2016). However, developing hearing-specific scales that consider the listening experiences of individuals...
with hearing impairment might be preferable because generic scales may not be sensitive to the experience of fatigue in individuals with hearing impairment, e.g. Hornsby and Kipp (2016).

Our findings indicate that that audibility might not be the only factor that contributes to hearing impaired individuals’ experience of listening effort and fatigue as suggested by the following outcomes: i) three groups of participants with different hearing levels (hearing aid users, cochlear implants user, and individuals with single-sided deafness) reported similar levels of listening effort and fatigue, ii) no correlation was identified between hearing levels and both listening effort and fatigue, i.e. more severe hearing impairment did not necessarily indicate increased listening effort and fatigue, and iii) participants with hearing impairment reported increased listening effort and fatigue compared to controls, despite using hearing aids or cochlear implants.

Currently, the main focus of hearing rehabilitation is on improving audibility by providing individuals with listening devices such as hearing aids or cochlear implants. It is appealing to assume that restoring/improving audibility will reduce or eliminate listening difficulties. However, individuals with hearing impairment commonly report having to exert a lot of effort to communicate despite the sounds being audible when using a hearing aid or a cochlear implant. Individuals with hearing impairment also report feeling tired and are sometimes tempted to remove their hearing aid at the end of the day as a result of having to concentrate for prolonged periods of time. Our findings suggest that improving audibility via the use of hearing aids or cochlear implants will not necessarily eliminate the experience of listening effort and fatigue. In addition, previous research
suggests that certain hearing aid features or rehabilitative techniques might not result in measured improvement in terms of audibility but could still reduce individuals’ experience of listening effort. For example, Sarampalis et al. (2009) did not report an improvement in the performance of a speech-in-noise task when participants with hearing impairment used a noise reduction algorithm. However, participants’ performance on a dual task paradigm improved significantly, which was interpreted as decreased listening effort.

The absence of correlation between hearing level and both listening effort and fatigue highlights the importance of identifying factors that contribute to listening effort and fatigue and that can be targeted in hearing rehabilitation. The findings of Study Two suggest that perceived hearing handicap correlates with both listening effort and fatigue. The correlation between hearing handicap and listening effort and fatigue is consistent with Motivational Control Model [MCM, (Hockey 2013)] and with the Framework for Understanding Effortful Listening [FUEL, (Pichora-Fuller et al. 2016)]. Both MCM and FUEL suggest that exerting cognitive effort does not necessarily translate into the experience of effort and fatigue as long as increased effort is rewarded by perceived successful performance (i.e. effective effort) or if a person is highly motivated to engage in the task. In fact, the distinction between “perceived effort” and “processing effort” has been highlighted in recent research (e.g. Lemke and Besser 2016) considering that the exertion of cognitive resources during listening does not always translate into a self-reported state of listening effort or fatigue. The association between aspects of motivation and perceived difficulties with listening effort and fatigue has important implications for hearing rehabilitations. Results of Study Two suggest that considering psychological
factors in hearing rehabilitation is likely to improve individuals’ benefit. Results of Study Two also suggest that the reward associated with being able to communicate and interact with others is likely to alleviate the negative experiences of effort and fatigue. Also, the motivation to engage in different listening situations is unlikely to be associated with perceived effort or fatigue if the reward associated with successful task performance justifies the expenditure of cognitive resources.

The assessment of self-reported listening effort and fatigue is of potential importance in clinical settings as it provides insight to individuals’ perception of these listening difficulties. However, unlike physiological measures, self-report measures do not provide information about the underlying mechanisms of listening effort and fatigue. Insight into the mechanisms of listening and fatigue may be critical to explaining the contradictory findings reported in the literature and inform effective strategies to reduce effort and fatigue. Behavioural and physiological measures of listening effort and fatigue may also provide an objective assessment of the benefit from using listening devices and rehabilitative strategies. Although any behaviourally or physiologically measured benefit might be of minimal importance if not subjectively reported, being able to “put a number” on how much improvement the patient gains would provide a useful index for the assessment of benefit. For example, if small gains are measurable with behavioural or physiological methods that are not detectable with self-report measures.

The difficulty in identifying a single reliable and valid measure for the assessment of listening effort and fatigue was constantly highlighted throughout this thesis. The absence of correlation between the measures and the inconsistent findings reported in the
literature raised questions about the reliability of the measures and about the underlying dimensions assessed by each of them. The findings of Study Three suggest that low reliability is not the reason for the lack of correlation between alternative measures of listening effort and fatigue. Simultaneous recordings of alternative measures of listening effort suggested that they tap into four separate underlying dimensions that are: i) “performance”: the ability to correctly respond to speech inputs, ii) “alertness”: being alert to the occurrence of a stimulus, iii) “Processing”: cognitive processing of auditory information, and iv) “behavioural consequences”.

The multidimensionality of the different measures suggests that using the measures interchangeably and reporting them as measuring the same underlying dimension might not be entirely correct. Our findings are in line with recent research that suggests that drawing conclusions based on the findings of a single measure of listening effort/fatigue can sometimes lead to inaccurate conclusions. Kramer et al. (2016) reported larger pupil size in participants with normal hearing compared to participants with hearing impairment, in the difficult listening condition only (in the presence of background noise that limited performance to 50% correct). Kramer and colleagues argued that larger pupil size could be interpreted as participants with normal hearing having to exert more listening effort than participants with hearing impairment if the interpretation did not consider aspects of performance and perceived effort. Kramer and colleagues suggested that increased pupil size in participants with normal hearing, along with better task performance and less self-reported effort compared to participants with hearing impairment, might instead be an indication of better ability to cope with the demands of the task as a result of having more cognitive resources, i.e. effective effort.
Comprehensive interpretation of the results obtained from different listening effort measures would benefit from considering at least the four dimensions identified in Study Three. Considering multiple dimensions would help in understanding why measures of listening effort do not always correlate with each other and would be useful for drawing conclusions on whether effort was effective or ineffective. Effort cannot be always considered a negative experience. Effective effort is generally associated with better ability to perform a task and does not necessarily translate into perceived effort. There are, of course, barriers to the ability to employ multiple measures in clinical or research settings. Therefore, before choosing an appropriate measure, it is essential to obtain an in-depth understanding and consideration of the aspect of increased listening demands to be assessed. Interchangeable use of the measures while assuming that they all tap into the general concept of listening effort might lead into inaccurate interpretations of the findings.
CHAPTER SEVEN

CONCLUSIONS AND FUTURE DIRECTIONS
CHAPTER SEVEN

CONCLUSIONS AND FUTURE DIRECTIONS

7.0 Overview

This chapter includes the conclusions of the three studies in this PhD thesis, in addition to suggestions for future research work.

7.1 Conclusions

Individuals with hearing impairment report increased levels of listening effort and fatigue in everyday life compared to individuals with good hearing. This finding highlights the potential importance of routine assessment of listening effort and fatigue as part of hearing disability (aka activity limitation in the current International Classification of Functioning Disability and Health; World Health Organisation 2001) in clinical settings. The assessment of listening effort and fatigue would provide information about difficulties experienced by hearing-impaired individuals that cannot be assessed using traditional audiological methods and current measures of hearing disability.

Listening effort and fatigue are independent of the severity of hearing impairment. Individuals with hearing impairment report increased levels of listening effort and fatigue despite using hearing aids or cochlear implants. Our findings suggest that a hearing rehabilitation strategy with a main focus on improving audibility, via the use of hearing aids or cochlear implants, is unlikely to eliminate the experience of listening effort and fatigue. However, considering psychological factors, such as motivation, in hearing rehabilitation might reduce patients’ experience of listening effort and fatigue. The experiences of listening effort and fatigue in individuals with hearing impairment are...
correlated with hearing handicap, which is consistent with the Motivation Control Model (Hockey 2013). Encouraging patients to recognise and focus on the pleasure and the positive experiences of listening may increase their motivation to engage in different listening situations and might reduce their perception of listening difficulties. In turn, this may increase the benefits obtained from using assistive listening devices which could ultimately reduce their experience of listening effort and fatigue.

Potential measures of listening effort may tap into different dimensions. Therefore, the interchangeable use of the measures is not to be encouraged. It is also important to consider that some of the measures might not be related to listening effort. For example, changes in pupil size might provide an indication of task engagement and this is not a direct measure of effort e.g. Ohlenforst et al (2017b). Although more work is required to identify how each dimension contributes to the experience of listening effort, our findings suggest that interpreting the experience of listening effort based on the findings of a single measure might result in inaccurate conclusions. Including multiple dimensions provides a comprehensive assessment of the experience of listening effort and will help in explaining why different measures do not correlate with each other or result in contradictory findings. For example, effective effort that results in perceived successful performance is less likely to result in high levels of self-reported effort. Potential limitations of using multiple measures in clinical or research settings highlights the importance of a comprehensive understanding of the dimension of listening effort of interest before choosing an appropriate measure.
The findings of Study Three provide a framework that has improved our understanding of the underlying dimensions of listening effort.

### 7.2 Future directions

- Future work should consider developing standardised, hearing-specific scales for the assessment of listening effort and fatigue in everyday listening situations. Hearing-specific scales that assess difficulties in specific listening situations are more likely to be sensitive to the increased demands imposed on individuals with hearing impairment than generic scales. Hearing-specific scales can also complement hearing rehabilitation by identifying and targeting the listening situations where individuals experience increased difficulties. The development of hearing-specific scales requires identifying challenging listening situations through focus groups of individuals with hearing impairment and audiologists before validating and establishing the reliability of the scale using the regular methods.

- Due to the absence of correlation identified between age and each of self-reported listening effort and fatigue, future work should consider investigating self-reported listening effort and fatigue in a group of participants with a wider age range. Evidence suggests that the experience of fatigue can increase with age (Avlund 2010). Recruiting participants with a wider age range would provide a more comprehensive method for identifying whether age has an influence of the correlation identified between handicap and self-reported fatigue and whether age actually contributes to the experience of fatigue in individuals with hearing impairment.

- The direction of relationship between aspects of listening effort, fatigue, and
Chapter Seven

hearing handicap could not be established due to the limitations of the correlational design used in this PhD thesis. Future work should consider evincing the direction of the established relationships in order to inform hearing rehabilitation. For example, if increased handicap is the cause of increased self-reported fatigue, then encouraging participants to identify the positive experience of listening and improving their motivation is likely to alleviate the experience of fatigue. Establishing the direction of the causal relationships can be achieved by experimentally manipulating the hypothesised causal factors, or through investigating the effects of intervention, such as investigating the effect of increasing patients’ motivation on perceived listening effort and fatigue.

- Adjusting SNR to achieve 71% criterion performance in Study Three might have minimised the differences in listening effort between participants with different hearing levels. The noise level might have been too challenging for some participants as it is not representative of noise levels that are likely to be encountered in real life (Smeds et al. 2015). Evidence suggests that performance is likely to influence the amount of listening effort exerted by individuals (Pichora-Fuller et al. 2016), e.g. a challenging listening situation can demotivate the exertion of listening effort as a result of poor performance. Individuals’ experience of listening difficulties in real-life situations is of interest in hearing rehabilitation. Therefore, future work should consider obtaining simultaneous recordings while participants perform a listening task in different levels of background noise. This approach will determine whether the present findings apply to listening situations that are more representative of real-life SNRs. Performing the listening task at different SNRs would also allow an in-depth investigation of the effect of task performance on listening effort. Understanding the effect of performance on listening effort can be achieved by investigating how: i) fluctuations in task demands affect participants’ motivation and ii)
task engagement influences the findings obtained using the different measures. Future work should also consider obtaining simultaneous recordings of the measures while participants perform a more cognitively demanding task in the presence of background noise that is likely to be encountered in real-life situations, e.g. multi-talker babble. In real-life situations, listening to speech involve processing and retention of information. Using a more cognitively demanding listening task that requires manipulating perceived information would provide methods for identifying whether the findings of Study Three apply to situations that provide a closer approximation of real-life situations.

- Including measures of cognitive ability should be considered in future work. Individual differences in cognitive abilities might influence the experience of listening effort. For example, it has been suggested that individuals with increased working memory capacity are likely to experience less listening effort (Rönnberg et al. 2013). Measures of cognitive abilities might help in explaining individual differences in listening effort and fatigue despite of similar hearing level and is likely to complement the framework proposed in Study Three. Identifying an influence of cognitive abilities on participants’ experience of listening effort and fatigue can inform intervention by considering different audiological strategies in the rehabilitation of individuals with different cognitive abilities.

- Future work should consider establishing the feasibility of using listening effort measures in clinical settings by considering aspects such as the time that would be required for testing and the training that needs to be provided for the audiologists. Future work should also consider establishing the sensitivity of the measures to increased listening effort at the individual level. The conclusion of most listening effort research is based on average scores across participants. However, measures need to be sensitive to
changes in listening effort/fatigue at the individual level before they are considered for use in clinical practice. If sufficiently sensitive measures are clinically feasible they will complement the assessment of disability and improve the benefit that patients gain from hearing rehabilitation.

- The conclusions of Study Three were based on the results of testing a group of older adults with hearing impairment. Future work should consider investigating whether the findings apply to other age groups of participants with hearing impairment. An effect of age on listening effort has been reported in the literature (Desjardins and Doherty 2013). The effect of age on listening effort might limit the ability to generalise the findings of Study Three to other age groups of individuals with hearing impairment. Identifying whether the findings apply to participants from other groups is essential provided that the focus of hearing rehabilitation can be influenced by the age of the individual with hearing impairment.
REFERENCES


References


### Table 1. A summary of recent research papers (between 2015 and 2017) that have used self-report, behavioural or physiological measures in the assessment of listening effort including a description of: the measure used, the listening task, the main findings, and the interpretation. The papers are listed in chronological order. RT: reaction time, SNR: signal to noise ratio, HL: hearing loss, WM: Working Memory.

<table>
<thead>
<tr>
<th>Publication</th>
<th>Measure</th>
<th>Participant</th>
<th>Listening task</th>
<th>Main finding</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pals et al. (2015)</td>
<td>Two dual task paradigms</td>
<td>Adults with normal hearing (18-25 years old)</td>
<td><strong>Speech material and task:</strong> Sentence repetition</td>
<td>• Decreased RT when listening to speech in noise compared to quiet&lt;br&gt;• Increased RT to auditory input (but not for visual input) at 79% intelligibility compared to near ceiling intelligibility</td>
<td>• RT is sensitive to effort associated with listening to unintelligible speech and to listening to speech in noise&lt;br&gt;• RT to auditory stimuli might be a candidate measure of effort in clinics</td>
</tr>
<tr>
<td></td>
<td>One primary task: Speech recognition</td>
<td></td>
<td><strong>Listening conditions:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Two different secondary tasks:</td>
<td></td>
<td>• Quiet&lt;br&gt;• Background steady state speech shaped noise&lt;br&gt;• Background 8 talker babble</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1- RT to visual task&lt;br&gt;2- RT to auditory</td>
<td></td>
<td><strong>Criterion performance:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>stimuli</td>
<td></td>
<td>• 79% intelligibility&lt;br&gt;• Near ceiling intelligibility</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seeman and Sims (2015)</td>
<td>Skin conductance&lt;br&gt;Heart rate variability</td>
<td>Adults with normal hearing (18-38 years old)</td>
<td><strong>Listening material and tasks:</strong></td>
<td>• Decreased heart rate variability with increased dichotic task complexity and increased noise is the speech in noise task&lt;br&gt;• Increased skin conductance</td>
<td>• Heart rate variability appeared to be the most sensitive to changes in effort resulting from increased task complexity and increased</td>
</tr>
<tr>
<td></td>
<td>Heart rate&lt;br&gt;Dual task</td>
<td></td>
<td><strong>Listening conditions:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(primary task: speech in noise, secondary</td>
<td></td>
<td>1- Dichotic listening task (digits)&lt;br&gt;2- Speech in noise task (sentences</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>task: visual letter)</td>
<td></td>
<td>repetition)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Winn et al. (2015) | Pupillometry | Adults with normal hearing (18-33 years old) | **Speech material and task:** Repetition of vocoded sentences  
**Listening conditions:**  
- Noise channel vocoder (4, 8, 16, 32 channel)  
- Vocoder that simulates cochlear implant front-end processing (7, 11, 16, 21 dB/octave) | Increased pupil dilation with increased signal degradation even when word identification accuracy was 100% | Degraded speech results in increased effort despite optimal intelligibility  
- Pupillometry might be sensitive to decreased effort associated with |
| Wisniewski et al. (2015) | • EEG theta power  
• Self-reported effort | **Speech material and task:**  
Sentences in speech shaped noise; word identification task  
**Listening conditions:**  
-12 to 12 dB SNR in 6 dB steps | • Increased theta power and self-reported effort with decreased SNR  
• Correlation between theta power and self-reported effort | • Theta power can be used as an index of listening effort |
|-------------------------|-------------------------------|--------------------------|-------------------------------------------------|--------------------------------------------------|
| Francis et al. (2016)   | • Skin conductance  
• Pulse rate  
• Pulse amplitude  
• Self-reported task demand, effort, frustration, and performance (from the NASA TLX)  
• Adults with normal hearing (20-32 years old) | **Speech material and task:**  
Sentences repetition  
**Listening conditions:**  
• Unmasked speech  
• Distorted speech  
• Background speech shaped noise at -8 SNR  
• Background two talker babble at -8 SNR | • Improved performance in the unmasked speech condition compared to other conditions  
• Decreased self-reported demands in the unmasked condition compared to other conditions  
• No difference in the performance of the listening task or in self-  
• Physiological measures are sensitive to an aspect of increased listening demands that is not picked up by self-report measures or by conventional measures of speech recognition |
| Holube et al. (2016) | - Self-reported effort, stress, and speech recognition  
- Electrodermal activity (EDA) | - Adults with normal hearing (18 to 28 years)  
- Adults with hearing impairment (52-85 years) | **Speech material and task:**  
Sentences repetition  
**Listening conditions:**  
Degradation:  
- stationary background noise  
- Reverberation  
**Criterion performance:** | - Significant differences in self-report measures between the easy and the difficult condition in  
- Self-report measures were more sensitive to increased listening demands than EDA in lab settings that reported effort between the masked and the distorted speech conditions  
- Increased skin conductance, pulse rate, and pulse amplitude in the masked compared to the unmasked and distorted speech conditions  
- Generally, no correlation between self-report and physiological measures |
## Appendix A

| Lewis et al. (2016) | Verbal RT | Experiment one: children with normal hearing (5-12 years) | Experiment two: children | Speech material and task: Repetition of consonants, words, and sentences | Listening conditions: -5, 0, and 5 dB SNR | Increased RT for correctly repeated words with increased background noise | Increased effort with increased background noise | No difference in effort between

- Easy condition (performance accuracy 100%)
- Difficult condition (performance accuracy 30-80%)
- No significant difference in EDA between the different conditions in both participant groups
- Significant correlation between EDA and self-report measures in normal hearing participants when listening to speech in background noise
- did not induce much stress
with mild bilateral hearing loss, children with unilateral hearing loss, children with normal hearing (8-12 years) | RT between groups of children with normal hearing, mild HL, or unilateral HL | children with different hearing sensitivity

| Richter (2016) | Pre-ejection period (PEP) | Adults (mean age: 23.8) | **Speech material and task:** Sine wave discrimination **Listening conditions:**
- Easy (discrimination of sine waves of distant frequencies)
- Difficult (discrimination of sine waves of close frequencies) | Increased PEP in high demand-high-success importance (rewarding) condition | Increased PEP indicates increased effort
- Rewards associated with successful performance motivate the exertion of effort
- PEP is sensitive to the effect of motivation on effort

|  |  |  | Motivation:
- High success importance (reward for correct answers)
- Low success importance (no reward for correct answers) |  |  |
<table>
<thead>
<tr>
<th>Study</th>
<th>Speech material and task:</th>
<th>Listening conditions:</th>
<th>Brackets:</th>
<th>Multimodal presentation of speech might have increased processing demands on older adults and eliminated the benefit of visual cues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sommers and Phelps (2016)</td>
<td>• Serial recall task, i.e. repetition of last two /three words in a list of continuously presented words</td>
<td>• Younger adults (mean age: 20.1, PTA at 0.5, 1, 2 kHz: 0.33)</td>
<td>• Improved performance in the AV condition compared to the AO condition in younger adults but not in older adults</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Younger adults (mean age: 70.2, PTA at 0.5, 1, 2 kHz: 17.1)</td>
<td></td>
<td>• Multimodal presentation of speech might have increased processing demands on older adults and eliminated the benefit of visual cues</td>
</tr>
<tr>
<td>Wu et al. (2016)</td>
<td>• Two dual task paradigms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>One primary task: Speech recognition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Two different secondary tasks:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Visual RT (easy)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Stroop test (difficult)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Speech material and task: Hearing in Noise Test (repetition of sentences in background noise)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Listening conditions:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Maximum RT between 30% and 50% criterion performance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Decreased RT at the least and the most challenging SNRs.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• The pattern of change in RT as a function of SNR was similar across the two secondary tasks and across</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Increased RT with increased noise indicates increased effort</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Decreased RT in most challenging conditions might indicate cognitive overload/disengagement i.e. prioritising the reward associated with performing the “easier”</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Task Type</td>
<td>Task Details</td>
<td>Speech Material and Task</td>
<td>Listening Conditions</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Brennan et al. (2017)</td>
<td>Verbal RT</td>
<td>- Children with mild to severe hearing loss (8-16 years)</td>
<td>Speech material and task: Amplified non-sense word repetition</td>
<td>Listening conditions:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Adults with mild to severe hearing loss (19-65 years)</td>
<td>Model used as a control for comparison</td>
<td></td>
</tr>
<tr>
<td>Degeest et al. (2017)</td>
<td>Dual task:</td>
<td>- Dual task:</td>
<td>Repetition of a series of 5 monosyllabic digits from 0 to 12</td>
<td>Listening conditions:</td>
</tr>
<tr>
<td></td>
<td>Primary task:</td>
<td>Primary task: speech recognition</td>
<td>Repetition of a series of 5 monosyllabic digits from 0 to 12</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Secondary task:</td>
<td>Secondary task: visual memory task</td>
<td>In quiet, +2, and -10 dB SNR</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Self-reported listening effort</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Hsu et al. (2017)</strong></td>
<td><strong>Marsella et al. (2017)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>---------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Word recognition</em></td>
<td><em>EEG alpha power</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>RT to word</em></td>
<td><em>EEG theta power</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>categorisation</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>task (e.g.</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>determine if</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>one of two</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>pictures is an</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>animal picture)</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Children with</em></td>
<td><em>Children with</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>normal hearing</em></td>
<td><em>unilateral</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>(7-12 years)</em></td>
<td><em>hearing loss</em> (8-16*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>years)*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Speech material and task:**

- *Listening to single words in noise*  
- *Perform RT task*  
- *Repeat the word* perceived before RT task performance

**Listening conditions:**
- *Quiet, 0, and -5 dB SNR.*  
- *High and low semantic complexity of the words in the categorisation task*

**Speech material and task:**

- *Forced-choice word identification (disyllabic words)*

**Listening conditions:**
- *Quiet (easiest)*

**between conditions**

- *Significantly improved secondary task performance in participants with no tinnitus*

**Poorer word recognition with decreased SNR**

**Decreased RT with increased semantic complexity and decreased SNR**

**In contrast word recognition, RT has increased sensitivity to the effort induced by increased semantic complexity of speech.**
**Appendix A**

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Metrics</th>
<th>Speech material and task</th>
<th>Listening conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binaural noise (4-talker babble background noise) Noise to the poorer ear Noise to the better ear (most challenging)</td>
<td>hearing loss in binaural noise and noise to worse ear conditions compared to the other listening conditions</td>
<td>- No significant difference in theta power across conditions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• No significant difference in theta power across conditions</td>
<td>• Decreased alpha power in the most challenging listening condition (noise to better ear) might be an indication of loss of attention and withdrawal from task performance</td>
<td></td>
</tr>
</tbody>
</table>

**Miles et al. (2017)**

- EEG alpha power
- Pupilometry
- Adults with normal hearing (22-34 years)

**Speech material and task:** Repetition of sentences in background 4 babble noise

**Listening conditions:**
- Signal degradation:
  - 16 channels vocoded
  - 6 channels

- Increased alpha power and pupil size at 16 compared to 6 channels vocoded sentences
- No difference in alpha power

- Pupillometry and EEG are sensitive to changes in speech resolution (i.e. how degraded the signal is)

- Pupillometry is...
| Picou et al. (2017a) | Three dual task paradigms  
One primary task: Speech recognition  
Three different secondary tasks:  
- RT to simple visual probe  
- RT to complex probe  
- Word categorisation (increased) | Adults (22-32 years)  
Children (9-17 years) | Speech material and task:  
Single words repetition  
Listening conditions:  
Close to 50% criterion performance | No effect of secondary task manipulation on the performance of the primary task in children and adults with normal hearing  
Worse performance in the word categorisation | In depth of processing might increase the sensitivity of the dual task paradigm to listening effort in adults only | 
|---|---|---|---|---|---|
| Picou et al. (2017b) | Dual task paradigm  
Primary task: word recognition  
Secondary task: word categorisation  
• Self-reported effort, tiredness, and control | Adults with normal hearing (22-32 years)  
Children with normal hearing (9-17 years) | Speech material and task:  
Single word repetition  
**Listening conditions:**  
Microphone settings:  
• Omnidirectional  
• Fixed directional  
• Bilateral beamformer  
Reverberation:  
• Low  
• Moderate  
Background noise:  
• 7 dB SNR  
• 4 dB SNR | More self-reported effort when using the omnidirectional microphone  
Improved secondary task performance at 7 dB SNR  
Improved secondary task performance when using directional or beamformer microphones at moderate levels of reverberation only | Directional microphones might reduce listening effort at moderate levels of reverberation  
Participants’ willingness to have control over the listening situation might be related to their experience of effort |
| Picou and Ricketts (2017) | • Dual task paradigm  
Primary task: monosyllabic word recognition  
• Secondary task: RT to word categorisation task | • Adults with normal hearing (22-30 years) | **Speech material and task:**  
Single word repetition (after responding to the secondary task)  
**Listening conditions:**  
• Using bilateral directional microphones  
• Using bilateral omnidirectional | • Improved secondary task performance in the bilateral and unilateral directional microphone conditions  
• The use of directional microphones can improve speech recognition and reduce listening effort |
| van den Tillaart-Haverkate et al. (2017) | • Reaction time | • Adults with normal hearing (19-34 years) | **Speech material and task:** Digit triples (identification and arithmetic tasks used in Houben et al. (2013)) **Listening conditions:** Processing:  
• Unprocessed speech  
• Speech processed using ideal binary mask (IBM) noise reduction algorithm  
• Speech processed using minimum mean square error estimator (MMSE) noise reduction algorithm  
• Background noise: −5, 0, +5, and +∞ dB SNR | • Similar speech identification across different listening conditions  
• Decreased RT in the arithmetic task when listening to speech processed using noise reduction algorithms compared to unprocessed speech  
• Less self-reported effort in the IBM condition compared to the other two conditions. | • Arithmetic reaction time task can be used as an objective measure of the benefit obtained from using noise reduction algorithms in terms of reducing effort |

| Visentin | • RT to word | • Adults with normal hearing | **Speech material and task:** | • No difference in effort | • Increased effort |
| and Prodi (2017) | Identification task | Self-reported normal hearing (19-41 years) | Forced-choice word identification (disyllabic meaningful words) | Listening conditions:  
- Steady state speech shaped background noise  
- Fluctuating masker background noise | Speech recognition across the two different listening conditions  
- Increased RT and self-reported effort in fluctuating masker | when listening to speech in fluctuating noise compared to listening to speech in steady state speech shaped noise |
|------------------|---------------------|------------------------------------------|---------------------------------------------------------------|-----------------------------------------------------------------|-------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------|
| Ward et al. (2017) | Dual task paradigm  
Primary task: speech recognition  
Secondary task: visual monitoring (reaction time) | Younger adults with PTA < 20 at the frequency range 0.25 to 8 kHz (18-24 years)  
Older adults with hearing levels ≤ 25 at octave frequencies 0.25 to 2 kHz and ≤ 45 at octave frequencies 4 to 8 kHz (56-82) | Speech material and task: Sentences repetition  
Listening conditions:  
- 4, 6, and 8 channel vocoded sentences | Significantly worse secondary task performance in older adults compared to younger adults. | Older adults experienced increased effort |
| Wisniewski | EEG theta | Adults with  
Speech material and task:  
Increased theta | Increased theta | Increased theta |
| (2017) | power  | self-reported normal hearing (19-34 years) | Auditory oddball paradigm (button press when hearing a high frequency tone interspersed among low frequency tones) **Listening conditions:**  
• Near in frequency (target tone= 515 Hz)  
• Far in frequency (target tone=1200 Hz) | power and Gamma band inter-trial phase coherence in the near compared to the far condition during active listening but not during passive listening | activity might indicate increased demands on the WM in the difficult condition  
• Increased gamma band inter-trial phase coherence might indicate increased attention required for stimulus encoding |

**References**


182


### Fatigue Assessment Scale

The following ten statements refer to how you usually feel on a daily basis. For each statement, choose the one answer that best describes how you feel on a typical day. Please give an answer to each statement, even if you do not have any complaints at the moment.

<table>
<thead>
<tr>
<th>Statement</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1- I am bothered by fatigue</td>
<td>Never</td>
<td>Sometimes</td>
<td>Regularly</td>
<td>Often</td>
<td>Always</td>
</tr>
<tr>
<td>2- I get tired very quickly</td>
<td>Never</td>
<td>Sometimes</td>
<td>Regularly</td>
<td>Often</td>
<td>Always</td>
</tr>
<tr>
<td>3- I do not do much during the day</td>
<td>Never</td>
<td>Sometimes</td>
<td>Regularly</td>
<td>Often</td>
<td>Always</td>
</tr>
<tr>
<td>4- I have enough energy for everyday life</td>
<td>Never</td>
<td>Sometimes</td>
<td>Regularly</td>
<td>Often</td>
<td>Always</td>
</tr>
<tr>
<td>5- Physically, I feel exhausted</td>
<td>Never</td>
<td>Sometimes</td>
<td>Regularly</td>
<td>Often</td>
<td>Always</td>
</tr>
<tr>
<td>6- I have problems starting things</td>
<td>Never</td>
<td>Sometimes</td>
<td>Regularly</td>
<td>Often</td>
<td>Always</td>
</tr>
<tr>
<td>7- I have problems thinking clearly</td>
<td>Never</td>
<td>Sometimes</td>
<td>Regularly</td>
<td>Often</td>
<td>Always</td>
</tr>
<tr>
<td>8- I have no desire to do anything</td>
<td>Never</td>
<td>Sometimes</td>
<td>Regularly</td>
<td>Often</td>
<td>Always</td>
</tr>
<tr>
<td>9- Mentally, I feel exhausted</td>
<td>Never</td>
<td>Sometimes</td>
<td>Regularly</td>
<td>Often</td>
<td>Always</td>
</tr>
<tr>
<td>10- When I am doing something, I can concentrate quite well</td>
<td>Never</td>
<td>Sometimes</td>
<td>Regularly</td>
<td>Often</td>
<td>Always</td>
</tr>
</tbody>
</table>
Appendix C

Listening Effort Assessment Scale

The following statements ask about the level of effort that you use when listening in daily life. On the line below each statement, please circle the number that best indicates how you usually feel.

1- Do you have to put in a lot of effort to hear what is being said in conversation with others?

Lots of effort

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

No effort

2- How much do you have to concentrate when listening to someone?

Concentrate hard

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

Not need to concentrate

3- How easily can you ignore other sounds when trying to listen to something?

Not easily ignore

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

Easily ignore

4- Do you have to put in a lot of effort to follow discussion in a class, a meeting or a lecture?

Lots of effort

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

No effort
5- Do you have to put in a lot of effort to follow the conversation in a noisy environment (e.g., in a restaurant, at family gatherings)?

<table>
<thead>
<tr>
<th>Lots of effort</th>
<th>No effort</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
<td></td>
</tr>
</tbody>
</table>

6- Do you have to put in a lot of effort to listen on the telephone?

<table>
<thead>
<tr>
<th>Lots of effort</th>
<th>No effort</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 5 6 7 8 9 10</td>
<td></td>
</tr>
</tbody>
</table>
Participant Information Sheet
Understanding disability: measuring listening effort and fatigue in people with hearing impairment (Phase 1; Self report questionnaires)

Researchers: Mrs. Sara Alhanbali, Professor Kevin J Munro, Dr Piers Dawes

Understanding disability: measuring listening effort and fatigue in people with hearing impairment
You are invited to take part in a research study. This research study is part of a PhD project undertaken by Mrs. Sara Alhanbali. Before you decide if you would like to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information.

What is the purpose of the study?
The aim of this study is to compare the effort and fatigue that is associated with listening in everyday life e.g., when talking to your family and friends or watching the TV.

Why have you been chosen?
We are looking for adults and young people with hearing loss as well as people with normal hearing.

What will I be asked to do if I agree to take part?
If you agree to take part in this study, we would like you to complete the enclosed effort and fatigue questionnaires. Completing both of them should not take more than 15 minutes. If you are an older adult with no known hearing problem, we will need to confirm your hearing is within normal limits. This involves having you listen to sounds with different loudness levels. You will be asked to press a button whenever you hear a tone. If the hearing test shows that you have a possible hearing problem, we will provide you with a copy of the test results and a covering letter to take to your GP.

What will happen when the study is complete?
The findings will improve our knowledge about the effort and fatigue associated with listening. We will present the findings at conferences attended by audiologists and publish the findings in the scientific literature. No identifying information will be included in any publication or presentation of the data. You will also receive a summary of the findings of the research if you wish.

Do I have to take part?
It is totally up to you to decide whether you take part or not. If you agree to take part, please do the following: 1) Sign the enclosed consent form. 2) Complete the questionnaires. 3) Complete the contact details return form if you are interested in finding out about further research that will be carried out at the University of Manchester. 4) Return all of the paperwork to us using the postage-paid envelope provided.

Will I be paid for participating in the research?
We will post a £2 voucher to you, at the address shown on your consent form, when we receive your completed questionnaires.
What are the possible benefits of taking part?
The information collected in this study will not benefit you directly, but may help to improve understanding of the problems that hearing impaired individuals face in everyday life.

What if something should go wrong?
If you have a concern about any aspect of this study, you should speak to the researcher who will do her best to answer your questions. If she is unable to resolve your concern or you wish to make a complaint regarding the study, please contact a University Research Practice and Governance Co-ordinator on 0161 2757583 or 0161 2758093 or by email to research.complaints@manchester.ac.uk

Will all information be kept confidential?
All research results will be kept anonymous. When results are reported it will not be possible to identify individual participants. Individuals from the University of Manchester, NHS Trust or regulatory authorities may need to look at the data collected during the research study to make sure it is being carried out appropriately. The individuals accessing the information have a duty of confidentiality to you as a research participant.

How will the confidentiality of the data be ensured?
The data obtained will be held for 3 years. Personal addresses, postcodes, email addresses and telephone numbers will be kept in order for us to provide you with feedback of the study results. All names and contact details will be stored in a password-protected file on a hard disk with restricted access within the School of Psychology at the University of Manchester. All participant data will be stored in a separate database to the database with names and contact details. Participants will be identified by code number for data storage.

Who has reviewed the study?
The study has been reviewed and approved by [15/SC/0113] Research Ethics Committee.

How are we going to obtain your consent?
If you wish to take part in the study, please sign and return the enclosed consent form. If you are aged 16 or under, you can agree to take part by completing the enclosed assent form. However, before you can be enrolled in the study, your parent or guardian also has to agree to you taking part by signing the consent form.

Where can I obtain further information if I need it?
If you require any further information before, during or after the study, please feel free to contact:

Mrs. Sara Alhanbali
A3.8 Ellen Wilkinson Building
University of Manchester
Oxford Road
Manchester
M13 9PL
Tel: 0161 275 8568
Email: sara.al-hanbali@postgrad.manchester.ac.uk
Appendix E

CONSENT FORM

Understanding disability: measuring listening effort and fatigue in people with hearing impairment (Phase 1; Self report questionnaires)

1. I confirm that I have read and understood the information sheet for the above named study.  

2. I have had the opportunity to consider the information, ask questions and where relevant I have had these answered satisfactorily.

3. I understand that my participation is voluntary and that I am free to withdraw my cooperation from the study at any time, without giving any reason and without affecting any of my legal rights or medical treatment.

4. If I do decide to withdraw, any data that had already collected with my consent would be retained and used in the study.

5. I understand that relevant sections of my medical notes and data collected during the study may be looked at by individuals from the University of Manchester, from regulatory authorities or from the NHS Trust, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records.

6. I understand that any research data may be shared or used in anonymous form e.g., in conference presentations, scientific articles, or in other research studies. My identity will remain confidential.

7. I agree to take part in the above study.

I agree to take part in the above project.

Name of participant ___________________________  Date _______________  Signature ______________________

Name of person taking consent ___________________________  Date _______________  Signature ______________________
Appendix F

HHIE

Hearing Handicap Inventory for the Elderly (HHIE)

Name: ___________________________ Date: __________________

The purpose of this scale is to identify the problems your hearing loss may be causing you. Check ‘Yes’, ‘Sometimes’, or ‘No’ for each question. Do not skip any questions. If you use a hearing aid, please answer the way you hear without a hearing aid.

<table>
<thead>
<tr>
<th>S-1. Does a hearing problem cause you to use the phone less often than you would like?</th>
<th>□ Yes (4) □ Sometimes (2) □ No (0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-2. Does a hearing problem cause you to feel embarrassed when meeting new people?</td>
<td>□ Yes (4) □ Sometimes (2) □ No (0)</td>
</tr>
<tr>
<td>S-3. Does a hearing problem cause you to avoid groups of people?</td>
<td>□ Yes (4) □ Sometimes (2) □ No (0)</td>
</tr>
<tr>
<td>E-4. Does a hearing problem make you irritable?</td>
<td>□ Yes (4) □ Sometimes (2) □ No (0)</td>
</tr>
<tr>
<td>E-5. Does a hearing problem cause you to feel frustrated when talking to members of your family?</td>
<td>□ Yes (4) □ Sometimes (2) □ No (0)</td>
</tr>
<tr>
<td>Question</td>
<td>Yes</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>-----</td>
</tr>
<tr>
<td>S-6. Does a hearing problem cause you difficulty when attending a party?</td>
<td></td>
</tr>
<tr>
<td>E-7. Does a hearing problem cause you to feel “stupid” or “dumb”?</td>
<td></td>
</tr>
<tr>
<td>S-8. Do you have difficulty hearing when someone speaks in a whisper?</td>
<td></td>
</tr>
<tr>
<td>E-9. Do you feel handicapped by a hearing problem?</td>
<td></td>
</tr>
<tr>
<td>S-10. Does a hearing problem cause you difficulty when visiting friends, relatives, or neighbors?</td>
<td></td>
</tr>
<tr>
<td>S-11. Does a hearing problem cause you to attend religious services less often than you would like?</td>
<td></td>
</tr>
<tr>
<td>E-12. Does a hearing problem cause you to be nervous?</td>
<td></td>
</tr>
<tr>
<td>S-13. Does a hearing problem cause you to visit friends, relatives, or neighbors less often than you would like?</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td><strong>E-14.</strong> Does a hearing problem cause you to have arguments with family members?</td>
<td>☐ Yes (4) ☐ Sometimes (2) ☐ No (0)</td>
</tr>
<tr>
<td><strong>S-15.</strong> Does a hearing problem cause you difficulty when listening to TV or radio?</td>
<td>☐ Yes (4) ☐ Sometimes (2) ☐ No (0)</td>
</tr>
<tr>
<td><strong>S-16.</strong> Does a hearing problem cause you to go shopping less often than you would like?</td>
<td>☐ Yes (4) ☐ Sometimes (2) ☐ No (0)</td>
</tr>
<tr>
<td><strong>E-17.</strong> Does any problem or difficulty with your hearing upset you at all?</td>
<td>☐ Yes (4) ☐ Sometimes (2) ☐ No (0)</td>
</tr>
<tr>
<td><strong>E-18.</strong> Does a hearing problem cause you to want to be by yourself?</td>
<td>☐ Yes (4) ☐ Sometimes (2) ☐ No (0)</td>
</tr>
<tr>
<td><strong>S-19.</strong> Does a hearing problem cause you to talk to family members less often than you would like?</td>
<td>☐ Yes (4) ☐ Sometimes (2) ☐ No (0)</td>
</tr>
<tr>
<td><strong>E-20.</strong> Do you feel that any difficulty with your hearing limits or hampers your personal or social life?</td>
<td>☐ Yes (4) ☐ Sometimes (2) ☐ No (0)</td>
</tr>
<tr>
<td><strong>S-21.</strong> Does a hearing problem cause you difficulty when in a restaurant with relatives or friends?</td>
<td>☐ Yes (4) ☐ Sometimes (2) ☐ No (0)</td>
</tr>
<tr>
<td>E-22. Does a hearing problem cause you to feel depressed?</td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>☐ Yes (4) ☐ Sometimes (2) ☐ No (0)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>S-23. Does a hearing problem cause you to listen to TV or radio less often than you would like?</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ Yes (4) ☐ Sometimes (2) ☐ No (0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E-24. Does a hearing problem cause you to feel uncomfortable when talking to friends?</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ Yes (4) ☐ Sometimes (2) ☐ No (0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E-25. Does a hearing problem cause you to feel left out when you are with a group of people?</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ Yes (4) ☐ Sometimes (2) ☐ No (0)</td>
</tr>
</tbody>
</table>
For clinician use only:

<table>
<thead>
<tr>
<th>Emotional (E) questions:</th>
<th>Determine presence of perceived emotional and situational hearing handicaps based on E and S scores.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 ___ 4 ___ 5 ___ 7 ___</td>
<td>0-16: No Handicap</td>
</tr>
<tr>
<td>9 ___ 12 ___ 14 ___ 17 ___</td>
<td>17-42: Mild to Moderate Handicap</td>
</tr>
<tr>
<td>18 ___ 20 ___ 22 ___</td>
<td>≥43: Significant Handicap</td>
</tr>
<tr>
<td>24 ___ 25 ___</td>
<td></td>
</tr>
<tr>
<td>Subtotal E: ______________</td>
<td>(52 maximum)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Situational (S) questions:</td>
<td></td>
</tr>
<tr>
<td>1 ___ 3 ___ 6 ___ 8 ___</td>
<td></td>
</tr>
<tr>
<td>10 ___ 11 ___ 13 ___</td>
<td></td>
</tr>
<tr>
<td>15 ___ 16 ___ 19 ___</td>
<td></td>
</tr>
<tr>
<td>21 ___ 23 ___</td>
<td></td>
</tr>
<tr>
<td>Subtotal S: ______________</td>
<td>(48 maximum)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Total score: ____________</td>
<td>(100 maximum)</td>
</tr>
</tbody>
</table>

If you found this form on the Internet and are going to use it, kindly become a fan of EAR Audiology, Inc. on Facebook. Visit www.earaudiology.com to learn about our services and products.

Participant information sheet (Study Two, Study Three)
Understanding disability: measuring listening effort and fatigue in people with hearing impairment (Phase 2; Lab based measures)

Researchers: Mrs. Sara Alhanbali, Professor Kevin J Munro, Dr Piers Dawes

Understanding disability: measuring listening effort and fatigue in people with hearing impairment
You are invited to take part in a research study. This research study is part of a PhD project undertaken by Mrs. Sara Alhanbali. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information.

What is the purpose of the study?
The aim of this study is to measure listening effort and fatigue. We would like to develop measures of listening effort and fatigue for use in education and health settings. Measures of listening effort and fatigue would be useful for understanding the difficulties experienced by hearing impaired individuals and in measuring the benefit of intervention.

Why have you been chosen?
We are looking for hearing impaired adults with different degrees of hearing loss.

What will I be asked to do if I agree to take part?
Testing will take place at The University of Manchester. Test appointments must be a particular day on a set schedule, although we can arrange them at a time of day to suit you. Each appointment will last around 1 hour. At the beginning of the session, the researcher will test your hearing. Then you would be asked to do three tasks; 1) identifying speech in background noise, 2) pressing a button as soon as you see a visual stimuli on a computer screen 3) rating listening effort and fatigue using two short pencil and paper scales. You will also be asked to provide a saliva sample (by chewing a cotton swab) at the beginning and the end of the session. While performing the first two tasks we will make some physiological recordings. To make the recordings, we would put sticky sensors on your hand and your forehead. We will also record the size of the pupil of your eye by having you look at a cross that appears on a computer screen. All testing is safe and should not cause any discomfort. The saliva samples will be analysed to identify the levels of stress hormones. These hormones indicate the levels of fatigue you experience. The samples will be disposed at the end of the study.

What will happen when the study is complete?
The findings will improve our knowledge about the effort and fatigue associated with listening. We will present the findings at conferences attended by audiologists and publish them in the scientific literature. No identifying information will be included in any publication or presentation of the data. You will also receive a summary of the findings of the research if you wish.
Do I have to take part?
It is your decision whether you take part or not. Even if you decide to take part but then change your mind, you can withdraw at any point in time, without giving any reason. This will not affect your normal clinical care. We will ask your permission to retain the data obtained until the point of your withdrawal in the consent form.

Will I be paid for participating in the research?
We will reimburse your travel expenses. We will also provide you with a £10 voucher.

What are the possible benefits of taking part?
The information collected in this study will not benefit you directly, but could help in understanding the listening effort and fatigue experienced by hearing impaired individuals. Listening effort and fatigue measures could be used to optimise hearing aid technology to reduce listening effort and fatigue for people with hearing impairment.

What if something should go wrong?
It is highly unlikely that you will be harmed in any way. If you have a concern about any aspect of this study, you should ask to speak to the researcher who will do her best to answer your questions. If she is unable to resolve your concern or you wish to make a complaint regarding the study, please contact a University Research Practice and Governance Co-ordinator on 0161 2757583 or 0161 2758093 or by email to research.complaints@manchester.ac.uk.

In the event that something does go wrong and you are harmed during the research you may have grounds for a legal action for compensation against the University of Manchester or NHS Trust but you may have to pay your legal costs. The normal NHS complaints mechanisms will still be available to you.

Will all information be kept confidential?
All research results will be kept anonymous. When results are reported it will not be possible to identify individual participants. Individuals from the University of Manchester, NHS Trust or regulatory authorities may need to look at the data collected during the research study to make sure it is being carried out appropriately. The individuals accessing the information have a duty of confidentiality to you as a research participant.

How will the confidentiality of the data be ensured?
The data of obtained will be held for 3 years. Personal addresses, postcodes, email addresses and telephone numbers would be kept in order to provide you with feedback of the study results. All names and contact details would be stored in a password-protected file on a hard disk with restricted access within the School of Psychology at the University of Manchester. All participant data would be stored in a separate database to names and contact details. Participants would be identified by code number for data storage.

Who has reviewed the study?
The study has been reviewed and approved by [ref: 15/SC/0113] Research Ethics Committee.

Where can I obtain further information if I need it?
If you require any further information before, during or after the study, please feel free to contact:
Mrs. Sara Alhanbali
A3.8 Ellen Wilkinson Building
University of Manchester, Oxford Road, Manchester M13 9PL
Tel: 0161 275 8568
CONSENT FORM

Understanding disability: measuring listening effort and fatigue in people with hearing impairment (Phase 2; Lab based measures)

Please tick the box where you agree

1. I confirm that I have read and understood the information sheet for the above named study.

2. I have had the opportunity to consider the information, ask questions and where relevant I have had these answered satisfactorily.

3. I understand that my participation is voluntary and that I am free to withdraw my cooperation from the study at any time, without giving any reason and without affecting any of my legal rights or medical treatment.

4. If I do decide to withdraw, any data that had already collected with my consent would be retained and used in the study.

5. I understand that relevant sections of my medical notes and data collected during the study may be looked at by individuals from the University of Manchester, from regulatory authorities or from the NHS Trust, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records.

6. I understand that any research data may be shared or used in anonymous form e.g., in conference presentations, scientific articles, or in other research studies. My identity will remain confidential.

7. I agree to take part in the above study.

I agree to take part in the above project.

__________________________________________  ____________________________  ____________________________
Name of participant                          Date                                      Signature

__________________________________________  ____________________________  ____________________________
Name of person taking consent                Date                                      Signature
**Appendix I**

**NASA TLX**

**Figure 8.6**

**NASA Task Load Index**

Hart and Staveland’s NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

<table>
<thead>
<tr>
<th>Name</th>
<th>Task</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mental Demand</td>
<td>How mentally demanding was the task?</td>
<td></td>
</tr>
<tr>
<td>Very Low</td>
<td>Very High</td>
<td></td>
</tr>
<tr>
<td>Physical Demand</td>
<td>How physically demanding was the task?</td>
<td></td>
</tr>
<tr>
<td>Very Low</td>
<td>Very High</td>
<td></td>
</tr>
<tr>
<td>Temporal Demand</td>
<td>How hurried or rushed was the pace of the task?</td>
<td></td>
</tr>
<tr>
<td>Very Low</td>
<td>Very High</td>
<td></td>
</tr>
<tr>
<td>Performance</td>
<td>How successful were you in accomplishing what you were asked to do?</td>
<td></td>
</tr>
<tr>
<td>Perfect</td>
<td>Failure</td>
<td></td>
</tr>
<tr>
<td>Effort</td>
<td>How hard did you have to work to accomplish your level of performance?</td>
<td></td>
</tr>
<tr>
<td>Very Low</td>
<td>Very High</td>
<td></td>
</tr>
<tr>
<td>Frustration</td>
<td>How insecure, discouraged, irritated, stressed, and annoyed were you?</td>
<td></td>
</tr>
<tr>
<td>Very Low</td>
<td>Very High</td>
<td></td>
</tr>
</tbody>
</table>
Appendix J

VAS-F

Visual Analogue Scale to Evaluate Fatigue Severity (VAS-F)

Purpose: The scale consists of 18 items relating to the subjective experience of fatigue. Each item asks respondents to place an “X” representing how they currently feel, along a visual analogue line that extends between two extremes (e.g., from “not at all tired” to “extremely tired”). In contrast to discrete, Likert-type scales, the VAS-F places fewer restrictions on the range of responses available to individuals. However, the benefits of a visual analogue scale may be offset by the frequent reluctance of individuals to use the highest and lowest extremes.

Population for Testing: The scale has been validated with adults aged 18–55 years.

Administration: A self-report, paper-and-pencil measure, the scale requires between 5 and 10 min for completion.

Reliability and Validity: Initial psychometric evaluations conducted by Lee and colleagues [1] have demonstrated a high internal reliability ranging from .94 to .96. Concurrent validity has been established with the Stanford Sleepiness Scale and the Profile of Mood States scale. Still, some have criticized the scale as ambiguous, suggesting that it is not sensitive to the distinction between fatigue and sleepiness [2].

Obtaining a Copy: A copy can be found in the original article published by developers [1].

Direct correspondence to:
Dr. K. A. Lee
N411Y, Box 0606
Dept. of Family Health Care Nursing,
University of California
San Francisco, CA 94143-0606

Scoring: Each line is 100 mm in length – thus, scores fall between 0 and 100. The instrument also possesses two subscales: fatigue (items 1–5 and 11–18) and energy (items 6–10). Though individuals do not require training in order to score the scale, developers are quick to point out that high levels of inter-rater reliability are vital if results are to be correctly interpreted.

A. Shahid et al. (eds.), STOP, THAT and One Hundred Other Sleep Scales, DOI 10.1007/978-1-4419-9893-4_100, © Springer Science+Business Media, LLC 2012
Visual Analogue Scale to Evaluate Fatigue Severity (VAS-F)

ID # __________  Date __________  Time __________ a.m. __________ p.m.

We are trying to find out about your level of energy before and after your night of sleep. There are 18 items we would like you to respond to. This should take less than 1 minute of your time. Thank you.

DIRECTIONS: You are asked to circle a number on each of the following lines to indicate how you are feeling RIGHT NOW.

For example, suppose you have not eaten since yesterday. What number would you circle below?

not at all extremely
hungry 0 1 2 3 4 5 6 7 8 9 10 hungry

You would probably circle a number closer to the "extremely hungry" end of the line. This is where I put it:

not at all extremely
hungry 0 1 2 3 4 5 6 8 9 10 hungry

NOW PLEASE COMPLETE THE FOLLOWING ITEMS:

1. not at all  extremely
   tired 0 1 2 3 4 5 6 7 8 9 10 tired
2. not at all  extremely
   sleepy 0 1 2 3 4 5 6 7 8 9 10 sleepy
3. not at all  extremely
   drowsy 0 1 2 3 4 5 6 7 8 9 10 drowsy
4. not at all  extremely
   fatigued 0 1 2 3 4 5 6 7 8 9 10 fatigued
5. not at all  extremely
   worn out 0 1 2 3 4 5 6 7 8 9 10 worn out
6. not at all  extremely
   energetic 0 1 2 3 4 5 6 7 8 9 10 energetic
7. not at all  extremely
   active 0 1 2 3 4 5 6 7 8 9 10 active
8. not at all  extremely
   vigorous 0 1 2 3 4 5 6 7 8 9 10 vigorous
Appendix J

100  Visual Analogue Scale to Evaluate Fatigue Severity (VAS-F)

<p>| | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>9.</td>
<td>not at all efficient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td>not at all lively</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td>not at all bushed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td>not at all exhausted</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td>keeping my eyes open is no effort at all</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.</td>
<td>moving my body is no effort at all</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td>concentrating is no effort at all</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.</td>
<td>carrying on a conversation is no effort at all</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.</td>
<td>I have absolutely no desire to close my eyes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18.</td>
<td>I have absolutely no desire to lie down</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

extremely efficient
extremely lively
totaly bushed
totally exhausted
keeping my eyes open is a tremendous chore
moving my body is a tremendous chore
concentrating is a tremendous chore
carrying on a conversation is a tremendous chore
I have a tremendous desire to close my eyes
I have a tremendous desire to lie down

Copyright © KA Lee 1992. Reprinted with permission. Reproduction or duplication without written consent from Kathryn Lee is strictly prohibited.
Appendix K

Calculating pupil size in mm

The eye link 1000 calculates pupil size in pixels, i.e. an arbitrary unit which refers to the number of pixels in the camera image. Pupil size was changed into mm by calculating the number of pixels in an artificial pupil with known diameter.

A black circle with a known diameter was printed on a piece of paper. The paper was then taped to the chin rest in the position where a participant would rest his forehead. The distance between the chin rest and the camera was the same distance used when testing participants. A couple of trials where then run and the size of the artificial pupil was recorded. The recorded pupil size for the artificial pupil that had a diameter of 7mm was 3484 pixels. This means that the number of pixel points for an artificial pupil with an area of 38.5 mm\(^2\) = 3484 (provided that the area of the circle= \(\pi r^2\)). Consequently, pupil size in pixel was changes to mm\(^2\) based on the following equation=

Area in mm\(^2\) = (Area in pixel*38.5)/3484.