Real-world listening effort in adult cochlear implant users

Francisca Perea Perez, MSc

Thesis submitted to the University of Nottingham for the
degree of Doctor of Philosophy

March 2023

ABSTRACT

Cochlear implants (CI) are a treatment to provide a sense of hearing to individuals with severe-to-profound sensorineural hearing loss. Even when optimal levels of intelligibility are achieved after cochlear implantation, many CI users complain about the effort required to understand speech in everyday life contexts. This sustained mental exertion, commonly known as “listening effort”, could negatively affect their lives, especially regarding communication, participation, and long-term cognitive health.

This thesis aimed to evaluate the listening effort experienced by CI recipients in real-world sound scenarios. The research focused on social listening situations that are particularly common in everyday life such as having conversations in a busy café or communicating through video call. Additionally, some situations that prevailed during the COVID-19 pandemic were also examined (e.g., listening to someone who is wearing a facemask).

Multimodal measures of listening effort were employed throughout the research project to obtain a comprehensive assessment. Nonetheless, the primary focus was on measures that quantify objectively the cognitive demands of listening through a CI. To that end, we used a combination of physiological measures, functional near infrared spectroscopy (fNIRS) brain imaging and simultaneous pupillometry, both of which are compatible with CIs and capable of providing insights into the neural underpinnings of effortful listening. We also proposed a novel approach to quantify “listening efficiency”, an integrated behavioural measure that reflects both intelligibility and listening effort.

We successfully applied these assessments to 168 CI users and 75 age-matched normally hearing (NH) controls who were recruited throughout the project. We found that CI users experienced high levels of listening effort, even when their intelligibility was optimal under highly favourable listening conditions. Objective measures revealed that CI listeners exhibited significantly inferior listening efficiency than NH controls when listening to speech under moderate levels of cafeteria background noise and when attending online video calls. Physiologically, they showed elevated levels of arousal as revealed by larger and prolonged pupil dilations to baseline compared with NH controls, suggesting high cognitive load and increased need for recovery. The importance of visual cues was evident; the presence of video and captions benefited CI recipients by improving considerably their listening efficiency during online communication. These results were consistent with their subjective ratings of effort, both in the experiments and in daily life.

These findings provide objective evidence of the cognitive burden endured by CI listeners in everyday life. In addition, the objective assessments proposed were proved feasible to quantify the performance and cognitive demands of listening through a CI. In particular, listening efficiency showed sensitivity to differences in task demands and between groups, even when intelligibility remained near perfect. We argue that listening efficiency holds potential to become a CI outcome measure.

DECLARATION

I certify that this thesis is my own work, except where indicated by referencing. No part of this thesis has been submitted elsewhere for any other degree or qualification.

Francisca Perea Perez

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to all of the participants who kindly volunteered in this research. Special thanks go to all CI users that I met during the PhD and whose stories coping with hearing loss along with many other life's challenges really inspired and motivated me. This research would not have been possible without you.

Secondly, I would like to thank my supervisor Dr. Ian Wiggins for giving me the opportunity to complete this PhD, and for his continuous guidance, patience, and support throughout the project. I would also like to thank my co-supervisor Professor Douglas Hartley whose help and encouragement was key for me to complete this thesis. Thank you both for believing in me— you made me grow as a researcher!

My gratitude also extends to Dr. Pádraig Kitterick for his valuable contribution to this work as my third co-supervisor during a significant portion of my studies, and to Dr. Derek Hoare for reviewing my thesis and ensuring that I was supported in those stressful moments close to thesis submission.

I am indebted to Dr Ralph Holme from the Royal National Institute for Deaf People, to Ray Glover from the National Cochlear Implant Users Association, and to all organisations and charities that helped me distributing my studies and recruiting participants. Many thanks also to Becky, Sammi, Sandra, Vicenta, Adele, Beth and Bas for their contribution to the audio-visual material used in the online test, to Polly for her statistical advice, and to Guangting for revising my fNIRS analysis section.

Also thanks to all my PhD colleagues who were the best companions one could hope for on this PhD journey. Special thanks to Tia, Faizah and Yasmin (as part of *"the fabulous four"*) for the emotional support and joy that you brought me during the PhD. I am also grateful to Eleni, Steph and everyone at the "gym" for being such a wonderful group of colleagues and friends. I could not have made it without you.

Lastly but not least, my immense thanks go to my husband Seckin and my family whose endless love, care, and encouragement have always guided me through life and supported me in the very low moments. I have no words to express how lucky I am for having you in my life. *Gracias por estar a mi lado, ¡Os quiero muchísimo!*

TABLE OF CONTENTS

FINANCIAL SUPPORT

This PhD was supported by the Royal National Institute for Deaf People (RNID) and Cochlear Ltd. through a Ph.D. studentship (grant reference: S53). Additional funds were provided by the NIHR Nottingham Biomedical Research Centre (funding reference number BRC-1215-20003).

DISSEMINATION

CONFERENCE PRESENTATIONS

- Biomedical Research Centre Annual Conference, Nottingham (UK), November 2022, poster presentation.
- British Cochlear Implant Group (BCIG) Annual Meeting, Cardiff (UK), May 2022, poster presentation.
- N-trans Sandpit Meeting 3: funding focus, University of Nottingham, June 2021, oral presentation – 1st Prize “top ranked fellowship proposal and elevator pitch”.
- BCIG Virtual Meeting, online, May 2021, poster presentation.
- Sue Watson PGR presentation event, University of Nottingham, online, April 2021, oral presentation.
- Post Graduate (PGR) Student Research Meeting, University of Nottingham, online, March 2021, oral presentation.
- PGR Student Research Meeting, University of Nottingham, online, October 2020, oral presentation.
- BCIG Annual Meeting, Nottingham (UK), March 2020, oral presentation.
- Action on Hearing Loss (AoHL) Future Leaders Day, London (UK), February 2020, poster and oral presentation.
- ARCHES Conference, Paris (France), November 2019, poster presentation.

- Rovereto Attention Workshop Conference, Trento (Italy), October 2019, poster presentation.
- Medicine & Health Sciences Postgraduate Research Forum, Nottingham (UK), June 2019, poster and oral presentation.
- 5th International Conference on Cognitive Hearing Science for Communication (CHSCOM), Linkoping (Sweden), June 2019, poster presentation.
- Midlands Hearing Implant Programme Joint Team meeting, Birmingham (UK), May 2019, oral presentation.
- BCIG Annual Meeting, Southampton (UK), April 2019, oral presentation.
- Action on Hearing Loss (AoHL) Annual PhD Student Day, London (UK), January 2019, oral presentation.
- Neuroscience at Nottingham Conference, University of Nottingham, January 2019, poster presentation.

PUBLICATIONS

F. Perea Pérez, Douglas E.H. Hartley, Pádraig T. Kitterick, Adriana A. Zekveld, Graham Naylor, Ian M. Wiggins (2023). Listening efficiency in adult cochlear-implant users compared with normally-hearing controls at ecologically relevant signal-to-noise ratios. Frontiers in Human Neuroscience. https://doi.org/10.3389/fnhum.2023.1214485

F. Perea Pérez, Douglas E.H. Hartley, Pádraig T. Kitterick, Ian M. Wiggins (2022). Perceived Listening Difficulties of Adult Cochlear-Implant Users Under Measures Introduced to Combat the Spread of COVID-19. Trends in Hearing, 2022. https://doi.org/10.1177/23312165221087011

These publications were adapted to form chapters two and four, respectively.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACALES	Adaptive Scaling method for subjective listening effort
AB	Advanced Bionics
BA	Brodmann Area
BKB	Bamford Kowal Bench
BL	Baseline dilation
BOLD	Blood oxygen level dependent
BRC	Biomedical Research Centre
CBF	Cerebral blood flow
CI	Cochlear Implant
CMRO2	Cerebral metabolic rate of oxygen
CrI	Credible Interval
CSV	Comma Separated Value
DLPFC	Dorsolateral Prefrontal Cortex
EAS	Effort Assessment Scale
EEG	Electroencephalography
EFA	Exploratory Factor Analysis
ERPs	Event-related potentials
FAS	Fatigue Assessment Scale
FDR	False Discovery Rate
fMRI	Functional Magnetic Resonance Imaging
fNIRS	Functional Near-Infrared Spectroscopy
FP	Frontopolar
GLM	General Linear Model
GUI	Graphical User Interface
HA	Hearing Aid
HDI	Highest Density Interval
HHQ	Hearing Handicap Questionnaire
HI	Hearing Impaired
HL	Hearing Loss
HRV	Heart rate variability
HbO	Oxygenated haemoglobin
HbR	Deoxygenated haemoglobin
HbT	Total haemoglobin concentration
IFG	Inferior frontal gyrus
INT	Intelligibility
IPA	International Phonetic Alphabet
IQR	Inter-Quartile Range
LBA	Linear Ballistic Accumulator model
LC-NE	Locus Coeruleus- Norepinephrine
LE	Listening Effort
LIFG	Left Inferior Frontal Gyrus
LSTG	Left Superior Temporal Gyrus
MCMC	Markov Chain Monte Carlo algorithm

MEG	Magnetoencephalography
MFQ	Momentary Fatigue Questionnaire
MPD	Mean Pupil Dilation
NH	Normal Hearing
NIHR	National Institute for Health Research
NIR	Near-infrared light
NLP	Natural Language Processing
PET	Positron emission tomography
PFC	Prefrontal Cortex
PPD	Peak Pupil Dilation
PEP	Preejection period
RDLPFC	Right Dorsolateral Prefrontal Cortex
ROI	Region of Interest
RSpan	Reading span test
SATO	Speed–Accuracy Trade-Off
SCI	Scalp Coupling Index
SL	Slope to baseline
SNR	Signal-to-Noise Ratio
SSQ	Speech, Spatial and Qualities of Hearing Scale
STG	Superior Temporal Gyrus
STS	Superior Temporal Sulcus
TD	Task Disengagement
TFS	Temporal Fine Structure
TRT	Text Reception Threshold
UK	United Kingdom
USA	United States of America
WHO	World Health Organization
WM	Working Memory

CHAPTER 1. INTRODUCTION AND PROJECT DESCRIPTION

1.1 CHAPTER OVERVIEW

This chapter introduces the background of the project by providing an overview of the listening effort literature and its relevance in cochlear implant (CI) recipients. Specifically, it discusses the concept of listening effort (LE), theories of cognitive capacity, methods to measure LE, and previous neuroimaging studies investigating effortful listening in CI users. The chapter ends with the project aims and an overview of studies.

1.2 WHAT IS LISTENING EFFORT?

Although no consensus has been reached about a standard definition of LE, based on dictionary entries, Ronan McGarrigle and colleagues proposed a working definition that will be considered for the purpose of this thesis. They defined LE as "*the mental exertion required to attend to, and understand, an auditory message*" (McGarrigle et al., 2014). However, care should be taken when interpreting what an auditory message is, given that in complex auditory scenes, it often refers to speech perception but it can also refer to other auditory stimuli such as music or environmental sounds.

The study of LE has received increased attention across the hearing research community in recent decades. Since Kahneman wrote his book about attention and effort (Kahneman 1973), it became evident that listening involves interactions between auditory and cognitive systems (Pichora-Fuller et al. 2016). This complex interplay between bottom-up perceptual systems and top-down cognitive mechanisms is particularly relevant under adverse listening conditions. In fact, listening may become an effortful task when the brain needs to overcome and compensate for certain obstacles that could be present in: i) the acoustic source, e.g., degraded sound signal, accented speech; ii) the transmission path, such as noise, reverberation, electroacoustic signal processing limitations; or iii) the receptor or listener, e.g., hearing impairment limitations, cognitive decline, non-native speakers (Mattys et al. 2012).

While listening may be a relatively effortless process for the normal-hearing population, the presence of background noise can certainly increase the task demands, making listening more challenging (CHABA 1977). Indeed, listening in noisy environments requires several “backstage” brain operations to allow the selective attention needed to stay focused on a particular speech target while ignoring irrelevant competing sounds. This situation is quite common in real-world scenarios when multiple sound sources are present simultaneously such as in group conversations, the presence of background noise, etc.

These daily challenges are exacerbated when the listener also suffers from hearing impairment (HI). HI, deafness, or hearing loss (HL) refers to the total or partial inability to hear sounds. Under these circumstances, individuals deploy greater cognitive capacity to enable comprehension and memorisation of, and appropriate responding to, the perceived auditory message (Mattys and Wiget 2011; Rönnberg et al. 2013). Consequently, deaf and hard-of-hearing populations commonly complain about listening being a considerably taxing task (McGarrigle et al. 2014a; Pichora-Fuller et al. 2016). Such complaints were corroborated by previous research that provided evidence showing that HI listeners are considerably more affected by the presence of background noise than normally hearing (NH) individuals (CHABA 1977; Festen and Plomp 1990; Hygge et al. 1992; Larsby et al. 2005; Needleman and Crandellt 1995; Sarampalis et al. 2009). This debilitating effect of HL is consistent with the fact that HI individuals use sustained concentration and attention as coping strategies for hearing related issues. In other words, they selectively focus on the specific information (e.g., target speech) they want to hear while ignoring other sounds in the environment. However, the use of this strategy during sustained periods has the potential associated risk of experiencing feelings of stress, tension, and fatigue (Hetu et al., 1988).

Unfortunately, deploying increased cognitive resources with the effort that this entails does not always guarantee success in overcoming the difficulties of the listening task. Moreover, this mental exertion can occur even when individuals with

HL are able to understand and repeat every word of a sentence accurately, showing good intelligibility (Pals et al. 2020; Sarampalis et al. 2009; Winn, Edwards, and Litovsky 2015; Winn and Teece 2022).

This makes it difficult for audiologists to address these complaints given that the patient’s listening abilities assessed by speech-in-noise tests (word recognition scores) suggest good speech perception performance. This may lead to ineffective treatments since patients with similar scores may be provided with similar audiological solutions despite that their experiences may differ considerably in the cognitive load perceived. Hence, it is important to develop effective assessment and rehabilitation approaches that consider both auditory and cognitive factors.

In addition, there are other factors that can modulate how effortful a listening experience is perceived. For instance, the presence of visual cues during communication is known to aid speech recognition in CI users (Moberly, Vasil, and Ray 2020), which is likely to influence the LE experience. Nonetheless, it is not clear yet whether visual cues lead to reduced or increased LE (Picou, Ricketts, and Hornsby 2011, 2013). Likewise, other factors such as the degree of attention, motivational arousal, success importance, and pleasure can enhance or reduce the listener’s ability to concentrate and understand speech under adverse conditions. These modulating factors are discussed in the section 1.3.

1.3 FRAMEWORK FOR UNDERSTANDING EFFORTFUL LISTENING

The Fifth Eriksholm Workshop on “Hearing Impairment and Cognitive Energy” developed a heuristic theoretical model called “the Framework for Understanding Effortful Listening (FUEL)”. It adapted Kahneman’s Capacity Model of Attention (Kahneman 1973) to address specifically listening effort, while also considering other theories and models concerning cognition, motivation, and arousal (Figure 1.1).

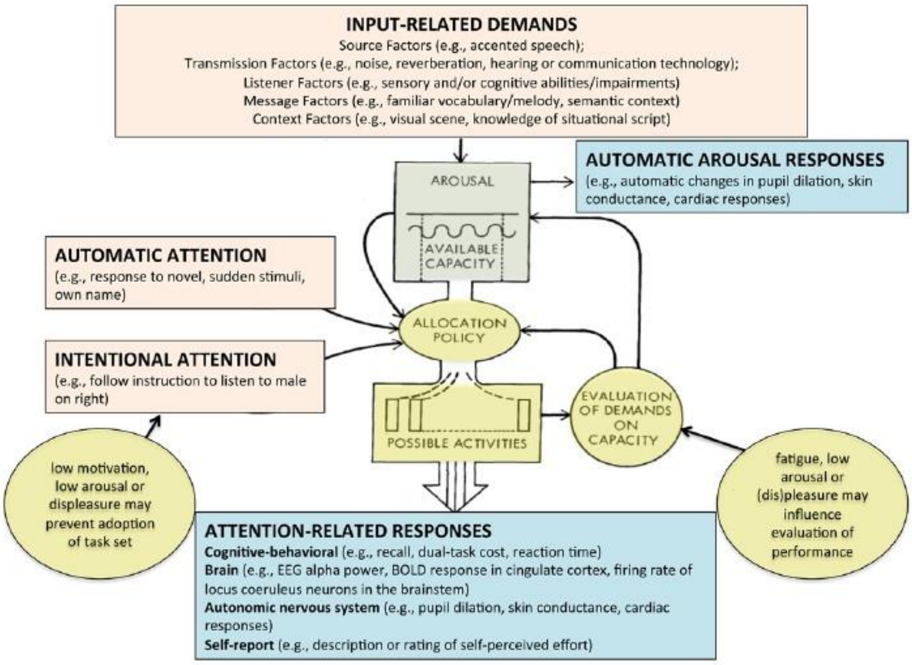


***Figure 1.1**. FUEL interpretation of Kahneman's (1973) Capacity Model for Attention in relation to listening effort and fatigue (Pichora-Fuller et al., 2016). Illustration reproduced by permission of Wolters Kluwer Health, Inc.*

The key principle of these cognitive theories is that humans have limited cognitive capacity available to be allocated in task performance (Kahneman, 1973; Cohen et al., 1994). These cognitive resources available are distributed for the execution of specific tasks according to the allocation policy or executive function (Kahneman 1973; Rudner 2016). Therefore, if increased attention and mental energy is focused on one task, a perceptual secondary task is more likely to show a reduced performance given that the pool of cognitive resources must be distributed between both tasks (Kahneman 1973; Lavie 1995). This idea is based on behavioural findings where dual-task paradigms have been used to investigate the resource models.

The way in which the available cognitive capacity is distributed depends on the task demands. In other words, how demanding the listening process is based on how many competing tasks need to be performed at the same time, and the listening adversities to be overcome (input-related demands). As mentioned before, these adversities are likely to be found in everyday listening contexts and include obstacles such as the presence of background noise, poor transmission in telephone or video calls, or the listener having HI. The deliberate allocation of mental resources to overcome these obstacles and proceed with a listening task is therefore defined in the FUEL model as LE (Pichora-Fuller et al. 2016).

Nonetheless, the allocation policy described by FUEL is not only modulated by the evaluation of task demands. Three other factors are believed to influence the resource allocation: first, the automatic attention, which is the intrinsic involuntary tendency to pay attention to sudden novel stimulus; second, the intentional attention that refers to the voluntary act of focusing attention and engaging in a task; and third, the effects of arousal. The latter comes from arousal theories that recognised the close correspondence between cognitive and arousal systems, given that variations of effort parallel variation of physiological arousal (Kahneman 1973).

The allocation policy ultimately yields the observed performance (attention-related responses), that can be measured by behavioural or self-report responses, brain activity, or by the autonomic consequences entailed in the actual task execution (Kahneman, 1973).

However, the most important contribution of the model proposed by the FUEL framework is the inclusion of fatigue, motivation, pleasure, and success importance as relevant factors able to modulate individuals’ willingness to attend and listen to an auditory message. The FUEL considered the motivational intensity theory that explains effort investment in goal pursuit based on a resource conservation principle (Brehm and Self 1989). Given the assumption that people’s cognitive energy is limited, individuals tend to save resources by investing just the amount of energy required for successful task execution. This theory also considers motivation as an important factor to influence effort mobilisation.

According to the motivational intensity theory, the motivation needed to engage in a listening task and maintain its execution depends on whether the listening goal has sufficient value to the listener, and on the importance of performing the task successfully (Richter, Gendolla, and Wright 2016). Regarding the latter, success importance is a factor that modulates an individual’s disposition to devote cognitive capacity in a given listening task. The idea is that more effort can be justified when performance has direct consequences for individuals' self-esteem or personal interests (Richter et al. 2016). As can be seen in Figure 1.2, when task difficulty increases low success importance leads to a lower effort mobilisation, whereas increased success importance allows higher effort exertion.

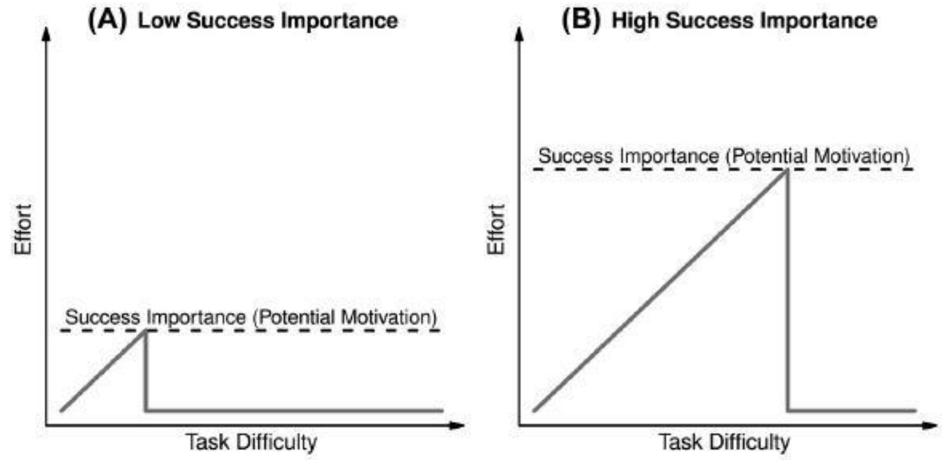


Figure 1.2*. Motivational intensity theory's predictions for task with fixed and known difficulty when low (A panel) and high (B panel) success importance influence individual's motivation (Richter et al. 2016). Illustration reproduced by permission of Elsevier.*

On the other hand, the value of listening can influence motivation, and thus effort exertion. Eckert et al. (2016) introduced the idea of neuroeconomics of listening as a conceptual and experimental framework for understanding listening effort. He suggested that neural systems are used during listening to support speech perception when the value from listening outweighs the relative cost of using these systems (Eckert, Teubner-Rhodes, and Vaden 2016). This value being, for instance, personal interest on achieving higher performance levels, receiving an overt reward (Kouneiher, Charron, and Koechlin 2009), and avoiding loss (Paulus et al. 2003). Figure 1.3 shows a prediction of how the value from listening would diminish with increasing listening difficulty in two different listening scenarios: a conversation with a loved one and listening to a documentary about lint.

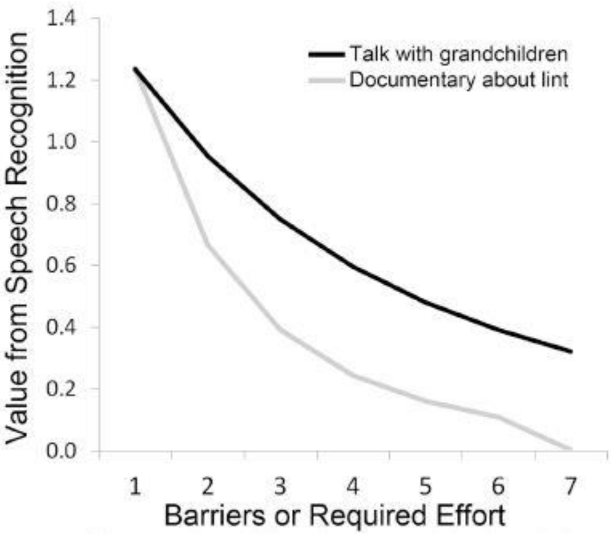


Figure 1.3. *Hypothetical listening effort discounting curves for speech content with different value to an older adult listener (Eckert et al.2016). Illustration reproduced by permission of Wolters Kluwer Health, Inc.*

As can be appreciated in the graph, the pleasure of talking to a loved one can act as a modulating factor of effort, which may keep the listener engaged even when the difficulty is high. In contrast, the curve is steeper in the less valuable scenario, which may lead to quicker disengagement. This interpretation is supported by Mohan Matthen (Matthen 2016) who also considered pleasure as a crucial factor to keep individuals with HL engaged despite listening barriers. He proposed that HI listeners should be taught to listen with their hearing devices in such a way that optimises pleasure.

Moreover, FUEL acknowledged low arousal, fatigue, and displeasure as factors that can reduce intentional attention and therefore, the likelihood of resource allocation. It is believed that sustained listening effort may induce listening related fatigue (a subclass of mental fatigue), since the continuous application of effort drains individuals' finite cognitive capacity (Hornsby, Naylor, and Bess 2016; Kahneman 1973). Given fatigue is associated with a reduction of cognitive processing abilities such as attention, concentration, processing speed, memory, and decision-making (Ackerman and Association 2011; Hetu et al. 1988; Kramer, Kapteyn, and Houtgast 2006), it is not surprising that it could lead to task disengagement (Boksem and Tops 2008; Hockey 2013). Indeed, previous research has provided evidence of the strong link between task disengagement and mental fatigue (Hopstaken, van der Linden, Bakker, and Michiel A. J. Kompier 2015).

Based on the effects of fatigue revealed by research studies, Richter et al. (2016) applied the prediction of motivational intensity theory as a function of individuals' ability and fatigue (Figure 1.4). According to this, highly fatigued people are likely to endure reduced levels of task difficulty than non-fatigued individuals at the same cognitive cost.

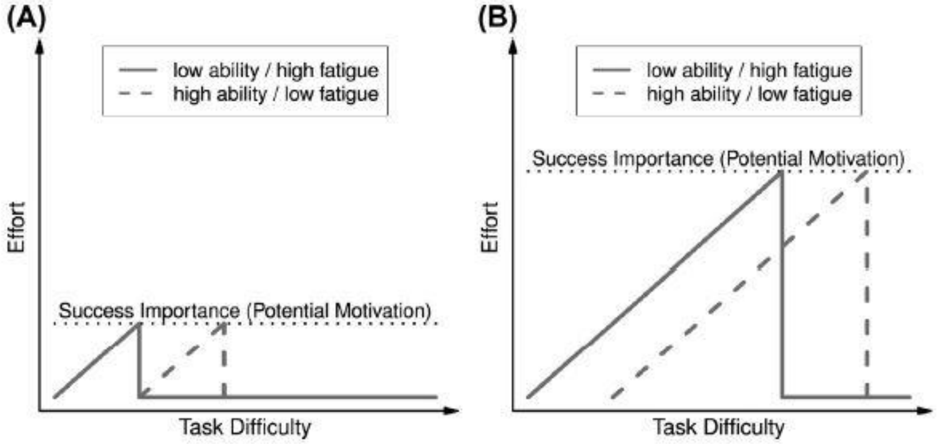


Figure 1.4*. Theoretical predictions for the impact of ability and fatigue on effort under conditions of known and fixed difficulty (Richter et al. 2016). (A) Low success importance. (B) High success importance. Illustration reproduced by permission of Elsevier.*

In summary, LE is a complex phenomenon that according to the FUEL model results not only from task demands and hearing difficulties but also from the motivation of individuals to expend mental effort in the challenging situations of everyday life. The model described the effort-motivation-demand relationship taking into account modulating factors such as fatigue, pleasure, and success importance. Therefore, FUEL constitutes a useful and robust framework for understanding listening effort that is based on well-established theories (such as motivational intensity, adaptive gain control, and optimal performance, fatigue, and pleasure), and supported by a vast foundation of research literature.

1.4 LISTENING EFFORT IN COCHLEAR IMPLANT USERS

1.4.1 Hearing devices as treatment for hearing loss

Hearing assistive devices such as hearing aids (HA) and cochlear implants (CI) are usually indicated as a treatment for hearing impairments. HAs are the most prevalent solution, commonly prescribed to listeners with mild or moderate degrees of HL. Their primary goal is to amplify the acoustic signal at certain frequencies to meet the specific needs of the user's HL. However, the use of HAs provides limited benefits to listeners with higher degrees of HL since their hair cells within the cochlea are usually damaged and cannot transmit the amplified acoustic signal to the auditory nerve. Under such circumstances, cochlear implantation is considered a more effective rehabilitation option. In the United Kingdom (UK), cochlear implantation is indicated when the patient meets the eligibility criteria defined in the National Institute for Health and Care Excellence guidance (NICE 2019).

These solutions, HAs and CIs, can also be combined to achieve what is called "bimodal hearing". CI users who have significant low-frequency residual hearing (HL≤100dB) in the non-implanted ear can benefit from a contralateral HA alongside the implant. Bimodal aiding can provide some of the benefits associated to binaural hearing (i.e., the ability to perceive sounds in both ears) such as sound localisation and improved speech intelligibility in noise conditions.

This thesis focuses on the LE experienced in relation to cochlear implantation. For a discussion about the effects of HAs and bimodal aiding on LE please refer to these articles: (Devocht et al. 2017; Hussein et al. 2022; Ohlenforst et al. 2017).

1.4.2 Cochlear implants (CI)

A CI is a treatment for severe-to-profound sensorineural hearing loss, where hair cells within the cochlea are often greatly damaged or completely absent. This medical device can partially restore hearing by directly stimulating the auditory nerve with electric pulses, and bypassing the inoperative hair cells. Although a CI effectively replaces the role of the hair cells, it does not restore normal hearing. Instead, the primary aim of cochlear implantation is to provide a useful representation of sounds in the environment, and in many cases, to enhance an individual's ability to discriminate speech.

The CI consists of external and internal components (Figure 1.5). The external component contains a microphone and a speech processor that picks up selected sounds from the environment and sends them to the internal component via the transmitter coil. The internal portion is comprised of the receiver, the stimulator, and the electrode array. The receiver, which is surgically implanted beneath the skin behind the ear, picks up the signal transmitted from the external component a via radio-frequency link. Then, the stimulator converts the sound signals into electrical pulses that are delivered through wires to the electrode array threaded into the scala tympani of the cochlea. The electrode array, which is interspersed along the cochlea, stimulates the auditory nerve with electric pulses. The auditory nerve transmits these neural impulses along the auditory pathway to the brain, where the impulses are perceived as sounds within the brain.

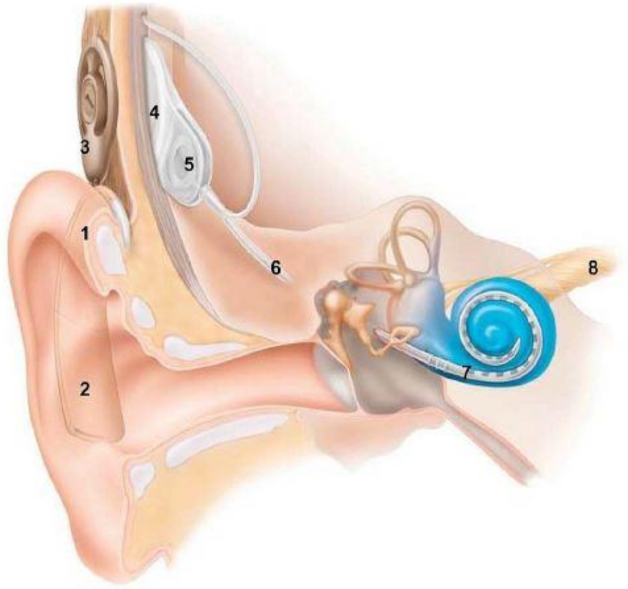


Figure 1.5. *Diagram of the internal and external components of a CI device. The numbers on the figure indicate: 1) the behind-the-ear speech processor with ear hook, 2) the battery case, 3) the transmitter coil, 4) the internal receiver, 5) the stimulator, 6) wires threaded into the cochlea, 7) the electrode array, and 8) the auditory nerve. Source: www.cochlear.com*

1.4.3 Is listening through a CI effortful?

Since their inception in the 1960s, the effectiveness of CIs have been widely proven especially in terms of speech recognition, but also regarding aspects such as overall quality of life (QOL), long-term well-being, and mental health (Boisvert et al. 2020; Buchman, Herzog, et al. 2020; Gaylor et al. 2013; Kitterick and Lucas 2016). However, this medical solution is not able to restore NH. Despite advances in CI technology, the quality of sound provided by a CI is still poor relative to the NH acoustic perception. This happens mainly due to the degradation or lack of certain auditory components of sound. Pitch, the perceptual correlate of frequency, and temporal fine structure (TFS) are probably the aspects of sound that are most profoundly affected by the implant (Caldwell, Jiam, and Limb 2017).

A typical CI array contains approximately 22 intracochlear electrodes that from apical to basal can transmit frequencies between approximately 200 to 8500 Hz. However, the NH cochlea uses 3,500 hair cells to transmit pitch, allowing for a frequency perception of 1,400 individual frequency steps between 20 to 20,000 Hz (Caldwell et al. 2017). This difference considerably reduces the pitch precision in CIs which is also worsened by channel interaction, when adjacent electrodes are simultaneously stimulated. As a result, CI users usually found it particularly difficult to distinguish acoustically similar consonant pairs such as /m/-/n/ and /t/-/k/(Rødvik et al. 2018).

Another important limitation of CIs is the absence of temporal fine structure (TFS) cues. Any sound can be separated mathematically into a slowly varying envelope and rapidly varying fine-structure component or carrier. The CI sound processor only codes information from the envelope, discarding the fine-structure component that better defines the periodicity of the signal. It is believed that the TFS information may be crucial for speech perception in the presence of noise, especially for fluctuating background sounds (Moore 2008). Therefore, the lack of TFS may partially explain the difficulty experienced by CI users when listening in background noise (Dincer D'Alessandro et al. 2018).

Other factors such as the compression in the acoustic dynamic range (Caldwell et al. 2017), i.e., the ratio between the loudest and softest sounds that an implant is able to transmit, the lack of binaural cues in the case of unilateral implantation, and individuals' biological limitations such as auditory nerve degradation and cochlear dead regions can further reduce CI performance.

Although these limitations do not prevent many CI users from achieving a satisfactory level of speech understanding, they could certainly hinder their ability to distinguish and segregate sounds, leading to the experience of LE. Indeed, previous studies using vocoders to simulate the reduced spectral resolution of auditory signals perceived by CI users have found that listening to degraded speech is associated with increased cognitive load, as revealed by behavioural responses (Başkent 2012), pupil dilation (Winn et al. 2015), and frontal brain activation (Lawrence et al. 2018; Wijayasiri, Hartley, and Wiggins 2017).

Despite the paucity of research testing LE in actual CI users, there is some evidence that also confirms this increased mental exertion. A few studies using dual-task paradigm showed that the degraded auditory input provided by CIs appears to induce LE even in optimal listening conditions (Perreau et al. 2017; Willis 2018). Likewise, recent neuroimaging studies observed increased activity in prefrontal brain areas of CI listeners compared to NH controls suggestive of greater cognitive processing required during speech perception (Dimitrijevic et al. 2019; Sherafati et al. 2021).

Overall, these studies suggest that CI users are likely to experience increased listening effort in both noisy and quiet environments, compared with NH individuals. These results are consistent with the perception of many CI listeners. Alhanbali and colleagues showed that CI users report high levels of LE and fatigue in everyday life (Alhanbali et al. 2017). Thus, improving our understanding of the LE phenomenon in the implanted population is crucial to develop appropriate assessments and interventions to ease the cognitive burden of listening through a CI.

Although it is unknown whether the high incidence of LE in the implanted population contributes to the high variance found in their speech perception outcomes (Hast et al. 2015), the sustained mental exertion that they report experiencing on a daily basis could have a negative impact on their lives. Certainly, the increased LE in HI individuals has been associated with communication difficulties (Hetu et al. 1988; Wie, Pripp, and Tvete 2010), social isolation (Kramer et al. 2006; Mick, Kawachi, and Lin 2014; Nachtegaal et al. 2009; Pronk et al. 2011; Shukla et al. 2020), and long-term challenges to cognitive health (Baltes and Lindenberger 1997; Lin et al. 2013; Peelle 2018; Rosemann and Thiel 2020) including a higher risk of dementia (Lin et al. 2011; Livingston et al. 2020). Therefore, the reduction of LE may be a key aspect of ensuring that CI users can prevent and/or overcome these challenges, improving their overall quality of life (Carlsson et al. 2015; Chia et al. 2007; Dalton et al. 2003).

1.5 HOW CAN LISTENING EFFORT BE MEASURED?

Although there is not a standardised measure of LE, different methods have been used in the literature to study this construct. Depending on the technique employed, the measures can be categorised into three main subgroups: i) self-reported, ii) behavioural, and iii) physiological measures (McGarrigle et al. 2014a; Pichora-Fuller et al. 2016).

1.5.1 Self-reported measures

Subjective measures of LE are designed to assess listeners' self-perceived effort. Such assessments typically consist of questionnaires in which listeners explicitly rate how much cognitive load they experience either while performing a listening task (measures of momentary cognitive load), or retrospectively in everyday life (questionnaires addressing daily-life listening experiences).

Self-reported measures of momentary effort are usually employed during experimental listening tasks and are provided immediately after each trial. They can use different scales to obtain participants' answers. Most commonly, visual analogue scales where listeners for instance have to rate from 1 to 10 how effortful a particular task was, or categorical scales where the degree of effort exerted is selected from different categories such as "no effort", "very little effort", "extreme effort", etc. The Adaptive Scaling method for subjective listening effort (ACALES) (Krueger et al. 2017) is an example of these measures. Likewise, the NASA Task Load Index (Hart and Staveland 1988) is another well-known and widely used measure for general task load that can be slightly modified for use in listening contexts (Francis et al. 2016).

In questionnaire-based ratings, individuals provide a retrospective judgement about the LE perceived in everyday life. These questionnaires assess the LE construct from a multidimensional perspective that encompasses several domains (such as speech understanding, spatial hearing, etc.) able to characterise the nature and extent of listening difficulty. Some examples are: the Effort Assessment Scale (EAS) (Alhanbali et al. 2017), the Speech, Spatial and Qualities of Hearing Scale (SSQ) (Gatehouse and Noble 2004) and its short version SSQ12 (Noble et al. 2013), the latter being more appropriate for use in time –sensitive contexts such as audiology clinics. Other relevant questionnaires that have been commonly used in the listening effort literature assess general fatigue (although not listening specific), and self-perceived hearing handicap. Some examples are the Fatigue Assessment Scale (FAS) (Michielsen et al. 2004) and the Hearing Handicap Questionnaire (HHQ) (Gatehouse and Noble 2004).

Subjective assessments of listening effort are particularly appealing since they are intuitive, quick and easy to deliver, and do not require particular expertise to administer and interpret (McGarrigle et al. 2014a). Moreover, they provide first-hand information about the effort perception, which makes them more sensitive than other objective methods when applied in audiological contexts (Johnson et al. 2015).

However, due to their subjective nature self-report measures are subject to individual bias. Indeed, people may rate LE differently either because they have different effort tolerance, for not being able to accurately recall past experiences (Francis and Love 2020), or for having different interpretations of effort (McGarrigle et al. 2014a). Recent evidence suggests that individuals may refer to their performance, or even how they feel, rather than the effort exerted when reporting their effort rating (Francis and Love 2020; McGarrigle et al. 2014a; Moore and Picou 2018). Therefore, although subjective measures accurately reflect people's perceived experiences, they should be interpreted with caution, especially when looking at inter-individual differences.

1.5.2 Behavioural measures

Behavioural measures of listening effort are objective assessments of the cognitive resources allocated to a listening task as demands increase but before the limits of available capacity are exceeded (Lunner et al. 2016). These measures can be divided into: single-task and multi-tasking paradigms (McGarrigle et al., 2014).

In the case of the single-task paradigm, participants are required to respond to a simple listening task by verbally identifying a speech stimulus (e.g., a heard word) or by pressing a response button (Gatehouse and Gordon 1990; Houben, van Doorn-Bierman, and Dreschler 2013). Participants' performance during the task, accuracy and response time (RT), are considered to reflect listening effort since both worsen (i.e., less accuracy and slower responses) as the level of task difficulty increases. Among them, RT is believed to better index LE since the speed in which participants provide their answers has been demonstrated to increase consistently as a function of task demands, even when intelligibility remained optimal (Houben et al. 2013).

However, absolute interpretations of RT cannot be made because it is not clear yet the relationship between the effort required to understand an auditory stimulus and the timing of providing the stimulus' response.

With regards to multi-tasking methodologies, dual-task paradigm is the most common behavioural technique applied in LE research to measure attention allocation (Gosselin and Gagné 2010; Karatekin 2004; McGarrigle et al. 2014a). It is in line with the resource model since it assumes that the brain has limited cognitive capacity. Therefore, dual-task experiments measure the reduction in performance, relative to a single-task baseline, that occurs when cognitive resources are distributed between two tasks. Then, the effort expended on the primary task can be measured by the performance on the secondary task, which is expected to be reduced as the primary task becomes more difficult. This deterioration in secondary task performance is then associated with increased cognitive load (Gosselin and Gagné 2010; Paas, Renkl, and Sweller 2003). Typically, dual-task experiments that aim to measure listening effort are composed by a primary speech recognition task that manipulates demands (e.g., in varying signal-to-noise ratios -SNRs-) and a secondary memory, tactile, or visual recognition task (McGarrigle et al. 2014a). Examples of these assessments are the Sentence-final Word Identification and Recall test, and the Cognitive Spare Capacity test, both involving a primary listening task and a secondary memory recall task (Rudner 2016).

Three cognitive domains are assumed to be involved when performing dual-task experiments: attention, working memory (WM), and processing speed (Pichora-Fuller et al. 2016). These cognitive domains are interrelated and are believed to represent the cognitive resources (also referred to attentional resources, processing resources, etc.) that are consumed during behavioural tasks.

Some behavioural tests are developed specifically for the assessment of WM, such as the Reading Span (RSpan) and the Text Reception Threshold (TRT) tests. They measure the cognitive ability that allows storing short-term memory, manipulating, and processing information simultaneously (Baddeley 2012; Daneman and Carpenter 1980). Although they do not provide a direct measure of LE, they are based on dual-task paradigm and are considered reliable predictors of individual differences in ability to understand speech under adverse conditions (Rönnberg et al. 2013). Indeed, there is evidence that individuals with higher WM capacity are likely to cope better with more demanding listening conditions than those with lower WM capacity (Foo et al. 2007; Larsby et al. 2005; Lunner 2003; Pichora-Fuller and Singh 2006; Rudner et al. 2012). These tests are therefore used as cognitive assessments to explore individual differences in cognitive capacity and/or WM.

Overall, behavioural measures, in particular the dual-task paradigm, are considered an effective assessment of listening effort that seems to be relevant to real life environments (McGarrigle et al. 2014a). Indeed, behavioural measures have been proposed as suitable tests for evaluating listening effort even in clinical contexts (Gosselin and Gagné 2010; Rönnberg, Stenfelt, and Rudner 2011; Rudner 2016).

Nonetheless, they also present some limitations. Behavioural measures are based on the assumption that individuals allocate their entire cognitive capacity, which is distributed between both tasks (Paas et al. 2003). However, there is currently no independent way of measuring the resources allocated to each task, or whether residual capacity still remains unused (McGarrigle et al. 2014a). Moreover, the voluntary control of effort allocation is limited in scope. As described in the FUEL, involuntary attention can be drawn towards novel stimuli, even when it does not correspond to the primary task and thus, goes against instruction. Children, for instance, are more prone to such distractions, making behavioural measures less reliable than when used in adults.

1.5.3 Physiological measures

Physiological measures provide another objective assessment to reveal systematic physiological changes that occur during task performance. When these physiological changes arise during challenging listening conditions relative to less demanding conditions, then those changes can be attributed to increased LE (McGarrigle et al. 2014a; Pichora-Fuller et al. 2016). Physiological measures can be categorised as: measures of brain activity (central nervous system), and measures of the autonomic nervous system. It should be noted that such classification is a simplification that refers to where the measure is taken from. However, it is important to clarify that these measures can be influenced, directly or indirectly, by both central and autonomic responses.

Table 1.1 lists the different techniques that have been used to obtain systematic physiological changes that can be attributed to LE.

Table 1.1 *Physiological measures of listening effort*

Type of activity to be measured	**Techniques**
Autonomic nervous system	Heart rate variability and preejection period (PEP)
	Skin conductance
	Hormonal responses (endocrine biomarkers)
	Pupillometry
Brain activity (central nervous system)	Positron emission tomography (PET)
	Magnetoencephalography (MEG)
	Electroencephalography (EEG) - Event-related potentials (ERPs)
	Functional magnetic resonance imaging (fMRI)
	Functional near-infrared spectroscopy (fNIRS)

Measures of autonomic nervous system

The measures included in this category reflect the involuntary activity of both the parasympathetic and sympathetic nervous systems that respond to the level of arousal experienced during task performance (Kramer, Teunissen, and Zekveld 2016). Some of these measures are skin conductance, change in pupil size, cardiac responses, and hormonal responses.

Skin conductance, for instance, is a measure of the skin’s capacity to conduct an electrical current (Boucsein 2012). When the amount of moisture present on the surface of the skin is augmented, it reflects increased activity in the sympathetic nervous system. Likewise, endocrine biomarkers provide measures of several hormones that are involve in the stress body response that occur under adverse listening conditions.

An increase in auditory task difficulty may also result in variations of the cardiac responses. Measures of heart rate variability (HRV) quantify the amount of heart variation in both time and frequency domains. It has been revealed that the reduction of parasympathetic nervous system that occur under demanding listening tasks is reflected by a reduction of high-frequency heart rate variability (Mackersie and Calderon-Moultrie 2016). Likewise, the pre-ejection period (PEP), an indicator of changes in the myocardial contraction force, is used to assess myocardial sympathetic activity and effort (Pichora-Fuller et al. 2016; Richter et al. 2016).

Pupillometry is probably the most commonly used technique when investigating listening effort (Zekveld, Koelewijn, and Kramer 2018). It measures the pupil dilation over time that is considered to be an index of cognitive processing load (Kahneman 1973; Kramer et al. 2016; Pichora-Fuller et al. 2016). Since pupillometry was used in this research, a detailed description of this technique is provided in the next section.

1.5.3.1 Pupillometry

Pupil size is determined by a balancing act between two smooth muscles of the iris: the sphincter and dilator pupillae muscles. The dilator muscle, innervated by sympathetic neurons in the superior cervical ganglion, expands the opening when it contracts. The sphincter is a circular muscle, innervated by parasympathetic neurons in the ciliary ganglion, that constricts the pupil. Therefore, changes in pupil size results from the balance between the sympathetic and the parasympathetic nervous system. Factors such as illumination level, fatigue, and cognitive activity affect the relative contribution of these two systems (Steinhauer et al. 2004).

Several studies have established a causal link between locus coeruleus (LC) activity and pupil size (Joshi et al. 2016; Reimer et al. 2016). In fact, there is empirical evidence that the Locus coeruleus-norepinephrine (LC-NE) system exercises cognitive control to regulate task performance (Aston-Jones and Cohen 2005), and that pupil diameter can be used to index LC activity (Gilzenrat et al. 2010).

The LC is a nucleus located in the upper dorsolateral pons (Figure 1.6) where norepinephrine is produced. It sends norepinephrine (NE) diffuse projections throughout the central nervous system. The LC-NE system plays a central role in integrating sensory information to modulate arousal, attention, and stress responses (Benarroch 2018). Indeed, optimal levels of NE in prefrontal areas of the brain have been found to be important to the facilitation of attention-related tasks and higher cognitive functions ranging from motivation to working memory.

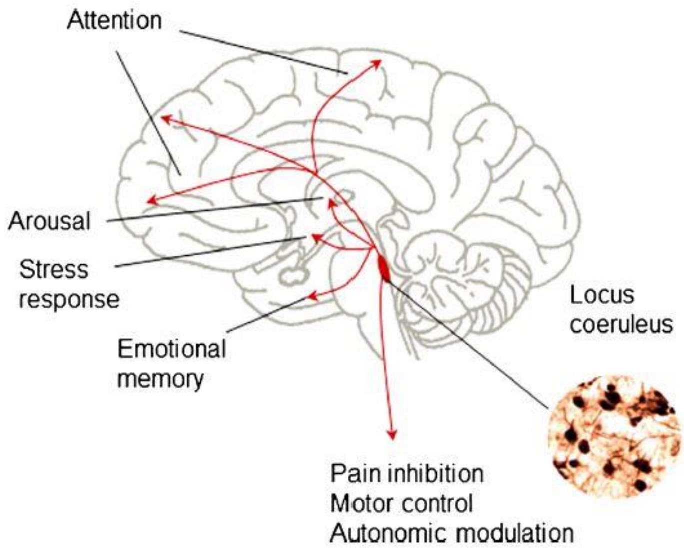


Figure 1.6. *Main effects of the Locus Coeruleus-norepinephrine (LC-NE) system (Benarroch 2018). Illustration reproduced by permission of Wolters Kluwer Health, Inc.*

The LC-NE system exhibits two main output modes: the phasic and the tonic mode (Aston-Jones and Cohen 2005). As shown in Figure 1.7, the phasic mode occurs during focused attention in response to task relevant stimuli and is characterised by intermediate baseline levels of NE. This mode is believed to support high task engagement, allowing accurate behavioural responses and filtering of irrelevant stimuli (Corbetta and Shulman 2002). Conversely, the tonic mode discharge sustained high levels of NE, which respond equally to task-relevant and task-irrelevant stimuli. This state is related to the arousal and waking state that happens during exploratory behaviour and it promotes distractibility. A third output mode is also considered where low levels of NE lead to diminished attention, disengagement from the task at hand, and low vigour in general (Aston-Jones and Cohen 2005).

These behavioural effects are similar to the ones typically observed during mental fatigue states (Hopstaken, van der Linden, Bakker, and Michiel A.J. Kompier 2015).

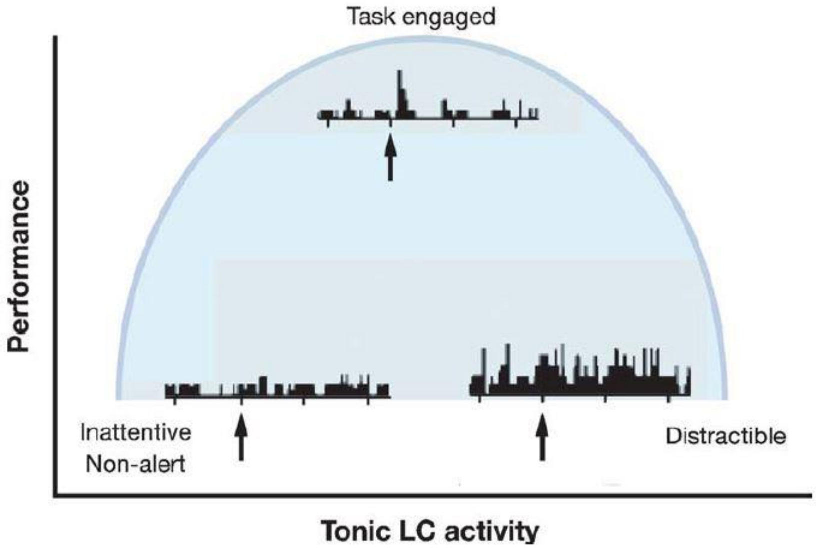


Figure 1.7*. Inverted-U relationship between Locus coeruleus (LC) activity and performance on tasks that require focused attention (Aston-Jones & Cohen, 2005). Illustration reproduced by permission of the authors and Annual Reviews.*

This inverted-U shaped relationship between performance and LC-NE activity is consistent with the classical Yerkes-Dodson law bell curve between performance and arousal. That pupil responses exhibit a similar inverted-U shaped relationship with task difficulty supports its suitability as an index of the LC cognitive control.

Indeed, pupillometry has become the most popular method to measure activation of the LC-NE system in response to cognitive processing load (Beatty 1982; Kahneman 1973). Many research studies have found that pupil dilation reliably changes with task difficulty following the same inverted-U shaped curve. Peak dilations occur at intermediate levels of difficulty, which may reflect the amount of effort or additional cognitive resources that a person is willing to exert to complete a task. On the other hand, pupil dilations are smaller for easy or very difficult tasks. In these cases, there is no need of extra cognitive control either because the person can complete the task easily or because the level of difficulty is so high that it becomes impossible or not worthwhile.

There is strong evidence of the sensitivity of pupil responses to mental workload, and particularly to listening effort (Naylor et al. 2018; Winn et al. 2018; Zekveld et al. 2018; Zekveld and Kramer 2014). Moreover, pupillometry meets the three sensitivity criteria proposed by Kahneman (1973). It shows sensitivity to within-task and between-task variations in processing load, as well as to individual differences in processing load.

Pupillometry system and data interpretation

Pupillometry systems use remote or head-mounted eye-trackers, a video camera, and mathematical algorithms to calculate the pupil's size and gaze location. It works as follows; an infrared illuminator creates reflection patterns on the cornea and pupil, which are captured and recorded by a video camera. The images of the subject's eyes and the light reflection patterns are then used to calculate the pupil diameter. Measures such as the mean and maximum dilation are usually calculated but time-series analysis is also used to model pupil dilation over time. The typical values of pupil dilations that can be obtained as a result of cognitive tasks are in the order of 0.1 to 0.5 mm, depending on testing conditions and tasks (Winn et al. 2018).

Care should be taken when interpreting pupil data since different parameters index different mechanisms. The peak pupil dilation, for example, reflects task-evoked momentary load, whereas the resting pupil diameter (baseline) before and after the presentation of the stimulus indexes participants' state of arousal or task engagement (Aston-Jones and Cohen 2005; Hopstaken, van der Linden, Bakker, and Michiel A.J. Kompier 2015). When the baseline pupil diameter is large, it is thought to reflect an "explorative" or "alertness" mode in which the person is not engaged in the task yet but rather explores the environment (Gilzenrat et al. 2010; Zekveld, Kramer, and Festen 2011).

Advantages and disadvantages of pupillometry

Pupillometry is a non-invasive, acoustically quiet, and portable technique that is relatively inexpensive and easy to use. Pupil headsets are devices designed to be lightweight and unobtrusive, somewhat similar to wearing a pair of glasses. Its simple design and functionality allow pupillometry to be used in combination with other measures of listening effort. Indeed, pupil responses are usually recorded while participants perform a behavioural listening task. It can also be employed simultaneously with other physiological measures, such as skin conductance and brain imaging (Alhanbali et al. 2019; Zekveld et al. 2014). This is a very interesting approach since pupillometry can complement other measures of LE such as brain imaging by enhancing their sensitivity and overall temporal resolution. In fact, pupillometry is a highly sensitive measure (it complies with Kahneman (1973)'s sensitivity criteria) that responds to task demands within seconds (temporal resolution of 60 to 1,000 Hz), allowing rapid changes in cognitive load to be captured.

Moreover, pupillometry is a time-series measurement that can be employed for the entire duration of an experiment. In this way, researchers can assess participants' cognitive effort at different time landmarks (e.g., before stimulus onset, during stimulus presentation, response processing and performance, etc.). Furthermore, since pupillometry is an optical based measure, its infrared technology does not interfere with CI devices. Therefore, it is a suitable technique to assess listening effort in CI users.

However, this technique also presents important limitations. Indeed, many factors besides cognitive effort can affect pupil responses. The most important one is luminance, able to evoke a maximum increase of 3 to 4mm in pupil size when changing from light to dark environments (Laeng, Sirois, and Gredebäck 2012). These changes are considerably larger than those evoked by cognitive tasks (0.1 to 0.5mm). With a lesser impact, other factors have also been found to affect the pupil response, including auditory stimulus characteristics (e.g., sound level), background noise, medication, emotions (Dionisio et al. 2001), eye and mental health problems, muscle pain and fatigue (Wang et al. 2018). Therefore, experiments aiming to assess cognitive load need to keep all these factors under control.

In addition, due to individual differences, there is no consensus on what percentage of change in pupil size corresponds to a particular change in proportion of effort capacity (Winn et al. 2018). Moreover, people present different dynamic range of absolute pupil size. Indeed, many studies have observed smaller pupil dilations in listeners who are older, listeners with hearing impairment (Koelewijn, Versfeld, and Kramer 2017; Zekveld et al. 2018), and listeners with traumatic brain injury (Koelewijn, van Haastrecht, and Kramer 2018), despite elevated effort usually being reported by these populations. In those cases, reduced pupil dilation should not be interpreted merely as reduced effort, but potentially as lower cognitive capacity or perhaps as reduced ability to maintain engagement (Winn et al., 2018). These issues can be addressed during data analysing by performing baseline corrections, proportionalisation (percent change in pupil size), and dynamic range normalization (percent points of individuals' dynamic range) within participants. Regarding the latter, Tepring Piquado and colleagues proposed a dynamic range normalisation (Piquado, Isaacowitz, and Wingfield 2010) that accounts for any age-dependent (Winn et al. 2018) and hearing-related (Koelewijn et al. 2017; Zekveld et al. 2018) changes in pupil responsivity, which allows comparison of pupil data between individuals regardless of their age or hearing status. Furthermore, the selection of adequate difficulty levels of the experimental stimuli (in accordance with the target population), and additional assessments of individual cognitive capacity such as working memory allows a better interpretation of pupil response (Winn et al. 2018).

Finally, although pupil responses provide an objective measure of cognitive load, these results are not always in line with self-perceived effort. Indeed, no correlations are usually found between pupillary responses and subjective measures of listening effort (Shields et al. 2023; Wendt, Dau, and Hjortkjær 2016; Zekveld and Kramer 2014; Zekveld et al. 2011). A possible explanation is that people's perception not only considers the momentary cognitive load but also other psychological factors (emotions, engagement, personal interpretations, etc.) and accumulative effects (fatigue) of their experiences.

Measures of brain activity

Brain activity measures in LE research have included positron emission tomography (PET); functional magnetic resonance imaging (fMRI); magnetoencephalography (MEG); electroencephalography/ event-related potentials (EEG/ERPs); and more recently, functional near infra-red spectroscopy (fNIRS) (Harrison and Hartley 2019; McGarrigle et al. 2014a; Pichora-Fuller et al. 2016; Wiggins et al. 2016).

The logic behind these neuroimaging techniques is that neurons generate electrical signals as they become active during cognitive task processing. Certain neuroimaging modalities, such as EEG and MEG, measure this neural activation directly by recording the average electric and magnetic field potential at different regions of the scalp. Conversely, metabolic neuroimaging methods, such as fMRI, PET and fNIRS are indirect measures of neuronal activity (Saliba et al. 2016). In these cases, the neural activation provokes changes in neurons' metabolism such as increase in oxygen demand. Since greater oxygen delivery is needed in the brain region involved in the task, consequently it leads to an increased in oxygenated haemoglobin (HbO) and a decrease in deoxygenated haemoglobin (HbR) of cerebral blood flow (CBF) in that area. These physiological parameters are then measured by metabolic neuroimaging methods.

However, not all of these techniques are compatible with CI devices and therefore nor suitable for neuroimaging CI recipients. In this regards, the main benefits and limitations of these techniques are discussed below:

- Functional MRI is usually the preferred neuroimaging technology due to its high spatial resolution. However, its use in actual CI users is contraindicated due to safety concerns. Indeed, a CI neuroprosthesis may heat, induce a current, or become dislocated when exposed to electromagnetic fields (Azadarmaki et al. 2014). Although some new CI prostheses are MRI compatible, the artifact created around the CI magnet obscures the acquisition of any brain activity on the ipsilateral side of the head. Another important constraint is motion artifacts (Quaresima, Bisconti, and Ferrari 2012). Subjects are required to remain completely still while in the scanner which is a serious restraint for children and newborns. Moreover, the poor temporal resolution, the scanner noise that interferes in auditory research, and the relative cost are additional downsides of this neuroimaging technique.

- PET is a nuclear functional imaging technique that has been used previously in language and cognitive processing research on CI recipients (Naito et al. 2000). PET is fully compatible with CIs and it is tolerant to subtle movements (Crosson et al. 2010). However, its main disadvantage is the radiation exposure that, for safety reasons, limits the number of scans that one subject can undertake. Its temporal resolution is also limited to the order of tens of seconds.
- EEG and MEG can measure the electrophysiological response of neural activation. The main benefit provided by these techniques is a high temporal resolution in the millisecond range (Babiloni et al. 2009), a safety profile that makes it suitable for CI users (even with several successive assessments), and its subtle movement tolerance that allows testing infants. Nonetheless, the principal drawback of using these techniques is the poor spatial resolution, which is inferior to other imaging modalities. Moreover, the electrical artifacts produced in EEG recordings in combination with CIs can produce data corruption (Gilley, Sharma, and Dorman 2008). MEG instrumentation can also interacts with the internal magnet of most CI models, which contributes to the paucity of MEG studies involving CI users (Pantev et al. 2006).
- fNIRS is a relatively novel optically-based neuroimaging technique that overcomes many issues associated with other neuroimaging techniques that are described above. Most importantly, fNIRS is safe for repeated use with CI users of all ages (Saliba et al. 2016), and has demonstrated potential as a LE measurement within both the NH and the CI user populations (Lawrence et al. 2018; McKay et al. 2016; Sherafati et al. 2021; Wijayasiri et al. 2017). Nonetheless, this technique also presents some limitations, such as the

limited depth of penetration that allows cortical activity measurements only and the inferior spatial resolution compared to fMRI.

Since fNIRS is the neuroimaging technique employed in this research, a detailed description is provided in Section 1.5.3.2.

1.5.3.2 Functional near infrared spectroscopy (fNIRS)

The technology of near-infrared spectroscopy was first described in 1977 by Professor Fransis Jöbsis (Jobsis 1977). He realised that due to the high degree of brain tissue transparency to light in the near infrared (NIR) spectrum (650-1000nm), NIR light was able to penetrate a wide range of body tissues including bone. This relative transparency occurs because biological tissues preferentially absorb light in the visible spectrum. This means that NIR light can penetrate through superficial biological layers and sample deeper tissue structures (Saliba et al. 2016).

Body tissue sampling is possible thanks to another important capability of NIR light, the detection of changes in haemoglobin oxygenation. Haemoglobin is the main pigmented molecule located in small vessels of the microcirculation. It is present in clinically significant quantities and exhibits oxygenation-dependent absorption of NIR light (Delpy and Cope 1997). The two main haemoglobin chromophores, oxyhaemoglobin (HbO) and deoxyhaemoglobin (HbR), are characterised each by different absorption spectrum (wavelenghts) which allows differentiation of their light absorption. Therefore, cerebral functionality is investigated using changes in concentration of HbO and HbR molecules and their timing with stimuli (e.g., auditory input).

During fNIRS imaging, two different wavelengths of NIR light are emitted simultaneously into the head. The light is scattered, diffused, and able to penetrate up to 1.5 cm (Elwell and Cooper 2011) through the tissue (Figure 1.8). Using a nearby light detector, the returned portion of NIR light (the backscattered light) that is not absorbed, can be collected and used to calculate changes in light attenuation of each chromophore. Each source-detector separation represents a measuring channel. The source-detector distance is very important since it will determine the penetration depth of the NIR light. The penetration depth is approximately half of the source-detector separation (Patil et al. 2011) (Figure 1.8). Although increasing the source-detector distance deepens the NIR penetration, it also leads to a worse SNR. This happens because the probability of light absorption is higher, and thus less returned light would be received by the detector. To ensure a compromise between depth sensitivity and SNR, typical values of source-detector separation are 30-35mm for adult studies and 20-25mm for infants (Pinti et al. 2018). These values provide penetration depths of approximately 1.3 and 1 cm, respectively.

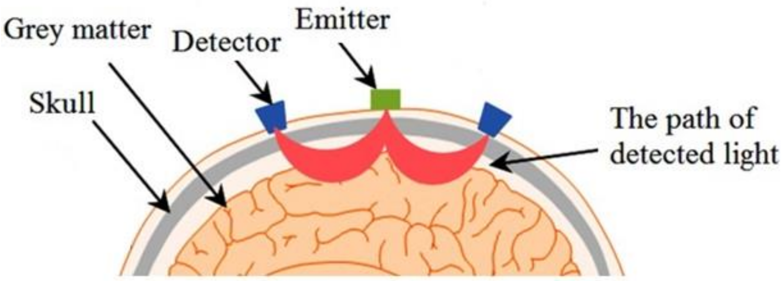


Figure 1.8. *Illustration of the near infrared (NIR) light path (shown in red) from source to detector through the different layers of the head. Illustration adapted from Naseer & Hong, (2015).*

When a brain area becomes active due to stimulation, the regional cerebral blood flow (CBF) increases immediately as the neural activity demands higher levels of oxygen. The changes in cerebral metabolic rate of oxygen (CMRO2) occur immediately after stimulus onset and are characterised by large increases in HbO and smaller decreases in HbR concentrations. This effect is often described as the washout effect (Wolf et al. 2002).

A typical cortical haemodynamic response reaches a peak at approximately 5 seconds from the stimulus onset and decays towards baseline approximately 16 seconds from the stimulus onset (Figure 1.9) (Mayhew et al. 2000). Similar haemodynamic responses have been found in the inferior frontal gyrus (Walsh et al. 2017) and auditory cortex (Zhang et al. 2018). However, the response’s amplitude, peak and undershoot latency duration can vary across different brain regions (Basura et al. 2018), task types, and participants' age (Pinti et al. 2018).

The relative difference in the concentration of oxygenated and deoxygenated blood between experimental conditions (or relative to rest state) allows the identification and localisation in the cortex of increased cerebral activity evoked by specific experimental tasks.

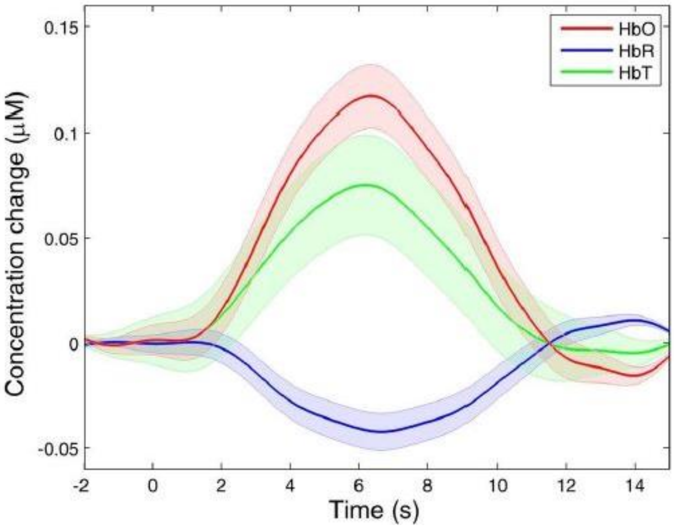


Figure 1.9. *Typical haemodynamic response measured by fNIRS. An increase in oxyhaemoglobin (HbO -red line) and a decrease in deoxyhaemoglobin (HbR- blue line) occur as oxygenated blood flows into the active brain region. HbT is the total haemoglobin concentration (green line). Source: https://www.gowerlabs.co.uk/fnirs*

Advantages and disadvantages of fNIRS

fNIRS has several advantages, compared with other neuroimaging techniques, that encourage its use in cognitive neuroscience and in particular in CI research. First of all, it is a non-invasive and safe neuroimaging technique that allows repeated measurements in close temporal intervals without presenting any safety concerns (Saliba et al. 2016). Due to its optical nature, this technology is fully compatible with CI devices, permitting the investigation of listening effort in the implanted population. Unlike other neuroimaging modalities such as fMRI, fNIRS is silent. Such quality is highly suitable for hearing research where auditory stimuli are usually presented. Moreover, fNIRS is reasonably robust to motion artifacts, allowing for various head positions and postures. Indeed, fNIRS has been used during walks (Herold et al. 2018), conversations (Suda et al. 2010), and even during dance movements (Noah et al. 2015). Such tolerance to movements is particularly useful for imaging paediatric populations. In addition, its portability allows patients to be scanned in comfortable environments, even clinical settings, with the option of changing locations with ease if necessary.

fNIRS has moderate spatial resolution, approximately 2-3 cm that allows a reasonable precision (better than EEG) to localise brain activation within specific cortical regions. A good temporal resolution commonly up to 10 hertz (which is higher than fMRI) permits a better track of the haemodynamic response (Pinti et al. 2018). Moreover, this technique provides a more comprehensive assessment of the haemodynamic response than fMRI, since both haemoglobin chromophores (HbO and HbR), as well as HbT, can be monitored simultaneously.

fNIRS is also feasible for multimodal imaging. It can be used in combination with other neuroimaging modalities such as EEG and fMRI, and with other physiological measures of listening effort, such as pupillometry. This would enhance its spatial and/or temporal resolution. Finally, fNIRS is an affordable technique with low running costs and no disposables, which places it among the most reasonably priced neuroimaging modalities, after EEG.

On the other hand, limited depth penetration is one of the main disadvantages of fNIRS. NIRS light is spatially limited to cortical layers up to 1.5 cm deep (Elwell and Cooper 2011; Ferrari and Quaresima 2012)(Elwell and Cooper, 2011). This is a considerable restriction for studies aiming to investigate deeper brain regions. As mentioned before, the cost of increasing the depth penetration is the worsening of SNRs and spatial resolution (Ferrari and Quaresima 2012). Moreover, although fNIRS' spatial resolution is better than that of EEG, it is considerably inferior to fMRI and thus, has no capacity for structural imaging. Nonetheless, the neuroanatomical localization of haemodynamic responses in the adult brain (Crosson et al. 2010) can be made using the International 10-20 positioning system (Jasper 1958).

Concerning data acquisition, hair is usually a nuisance because it can obstruct the contact between the optodes and the scalp, which is necessary to obtain good quality recordings. Moreover, participants with light colour hair are usually preferred since hair pigments can also attenuate the optical signal. Additionally, certain considerations must be taken into account when acquiring fNIRS data from CI users. Sometimes the external magnet or coil of the CI device can interfere with the fNIRS headset. In such circumstances, the headset should be placed over the magnet, which might obstruct the scalp contact of certain channels (Saliba et al. 2016). Nonetheless, the remaining channels should still provide useful data over a region of interest (ROI).

During data analysis, fNIRS recordings need to be filtered to eliminate physiological interferences. Indeed, other physiological responses such as heart rate and breath can interfere with the evoked cerebral responses of interest. A further caveat associated with data analysis is the lack of standardisation. To date, there are not standardised procedures or guidelines on the analysis and processing of fNIRS data as in other well-established technologies like fMRI (Pinti et al. 2018). This can lead to a disparity in procedures that makes direct comparisons between research studies and reproducibility of results difficult.

1.6 NEURAL CORRELATES OF EFFORTFUL LISTENING IN CI USERS

It has been observed, from neuroimaging studies of speech recognition, that engagement of non-auditory neural systems in challenging listening conditions supports performance and attention. Specifically, cingulo-operacular and fronto-parietal regions are suggested to be key in optimising performance in challenging listening conditions by providing adaptive control mechanisms. These mechanisms are related to intention and attention systems (Dosenbach et al. 2006). Nonetheless, the sustained engagement of these systems appears to have an effort-related cost, which is worth exerting just when the value from listening outweighs the relative cost (neuroeconomic) (Eckert et al. 2016).

Neuroimaging studies have consistently shown that intelligible sentences are processed by bilateral temporal cortex, complemented by inferior frontal gyrus (IFG) (Davis and Johnsrude 2003; McGettigan et al. 2012; Peelle 2018; Scott 2000). These areas, denominated in Figure 1.10 as the core speech network, are believed to form a hierarchy for intelligible speech processing (Peelle, Johnsrude, and Davis 2010). Conversely, lower levels of speech intelligibility (due to acoustically degraded speech) typically generate higher activity in cingulo-operacular, premotor, and frontoparietal cortex (Eckert et al. 2016; Peelle 2018; Wild et al. 2012). These areas would activate depending on the cognitive support required that may differ as a function of the acoustic challenge, and individuals' cognitive and listening abilities. The pattern of response in those regions tends to follow an inverted-U shaped response, so that greater activity occurs for degraded, yet still intelligible speech, than for completely clear or unintelligible speech. Regarding the latter, the low activity of those systems in very difficult listening conditions (Poldrack et al. 2001; Zekveld et al. 2006) is suggestive of task disengagement either for having exceeded participants' ability or simply because there is no value in keeping trying.

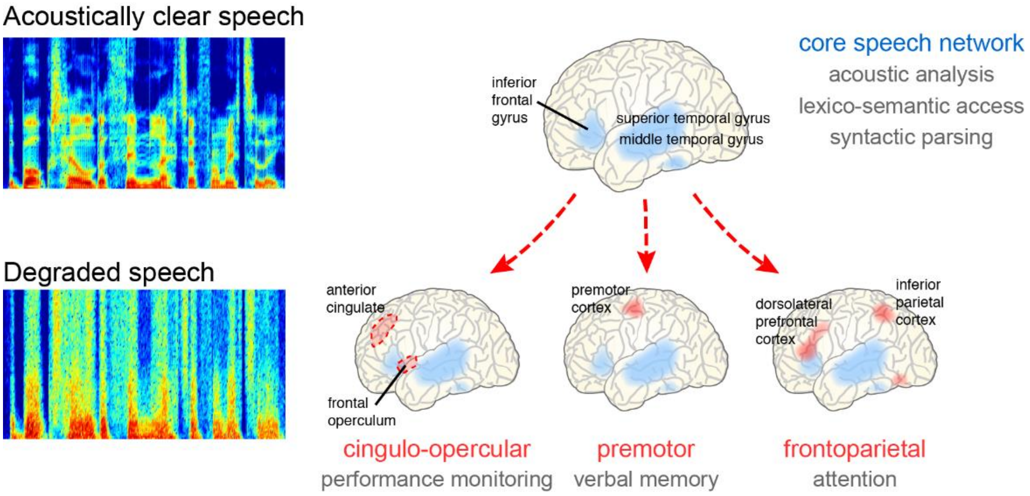


Figure 1.10. *Brain networks involved in processing clear and degraded speech (Peelle 2018). The core speech network is shaded in blue and is involved in both clear and degraded speech. The additional regions engaged to cope with the degraded speech signals are shaded in red. Illustration reproduced by permission of the author and Wolters Kluwer Health, Inc.*

The involvement of those brain areas in processing degraded speech, especially those belonging to attention-related subsystems such as the frontoparietal and cingulo-operacular networks, are consistent with the use of cognitive and executive processes required to support speech understanding during effortful listening (Eckert et al. 2016; Peelle 2018).

However, most of the available scientific evidence comes from neuroimaging studies testing NH listeners. Studies investigating effort-related neural activity in actual CI users are scarce. This paucity is mostly because implants are either contraindicated using traditional neuroimaging techniques such as fMRI, or render results difficult to interpret due to associated electro-magnetic artifacts (in the case of EEG measurements) (Wiggins et al. 2016). Nonetheless, there are some studies using CI simulations (e.g., vocoders) and a few on CI listeners that have investigated these neural networks involved during effortful listening.

PET neuroimaging literature investigating brain plasticity after cochlear implantation found that as post-lingually deaf adult CI recipients recover after implantation, they exhibit reduced activity in the superior temporal gyrus (STG) and increased activity in the left IFG (Strelnikov et al. 2015). Such effects occur because after implantation the compensatory visual activity in STG due to crossmodal reorganisation is gradually replaced by auditory-related activity. Likewise, increased activity in Brocas' area during speech processing is observed as CI patients become more experienced using their implants.

Nonetheless, these patterns of cortical activity may only apply to postlingually deaf CI recipients who achieved good intelligibility post-implantation (McKay et al. 2016; Petersen et al. 2013; Strelnikov et al. 2015). Indeed, poor speech perception after implantation has been associated with large and indistinguishable cortical activations, compared with brain activity in CI users with good speech perception (Olds et al. 2016).

Similar results were found using other neuroimaging modalities. An EEG study found that alpha oscillatory activity in the left inferior frontal gyrus (LIFG), canonical Broca's area, was positively correlated with self-reported listening effort in adult CI users

(Dimitrijevic et al. 2019). Likewise, there is evidence from EEG studies that even proficient CI users require the additional recruitment of prefrontal areas to support listening (Jiwani, Papsin, and Gordon 2016). Moreover, recent evidence from an optical imaging study revealed reduced auditory cortical activity (STG regions) and increased dorsolateral prefrontal cortex (DLPFC) activation in CI users compared to NH controls when listening to spoken words in quiet (Sherafati et al. 2022). The authors concluded that CI listeners experience greater cognitive load during speech understanding than NH listeners, as revealed by compensatory recruitment of the left prefrontal cortex (Sherafati et al. 2022).

These findings were also corroborated by studies using CI simulations in NH individuals, using 'vocoded' speech. Overall, frontotemporal activation was observed, with frontal regions (including LIFG) showing elevated responses to degraded speech, compared to clear speech (Lawrence et al. 2018; Wijayasiri et al. 2017; Wild et al. 2012; Zekveld et al. 2006, 2014). In Wijayasiri's study (2017), such prefrontal activation in LIFG was attention dependent suggesting that this pattern of activity was related to the effort associate with attentive listening rather than other aspects of the acoustic stimulus. Therefore, activation in LIFG was suggested as a potential neural marker of LE (Lawrence et al. 2018; Wijayasiri et al. 2017; Wild et al. 2012).

Based on the aforementioned evidence the neuroanatomical location of effort-related activity in CI users seems to involve prefrontal regions including the left inferior frontal gyrus (LIFG), positioned in Broca's area (Brodmann area [BA] 44 and BA45), and the DLPFC approximately located in BA9 and BA46.

It remains unknown, however, whether the same cortical patterns would be elicited when CI users listen to speech in realistic and more complex listening environments such as in the presence of background noise. In those situations, the cognitive processing that they already experience when listening to speech in quiet is likely to be exacerbated. In fact, it is well known the great susceptibility of CI users to noise exposure (Firszt et al. 2004; Fu and Nogaki 2005). However, to date there is no evidence on how the increased cognitive load of listening in real-world sound scenarios may be reflected in the brains of CI listeners.

It is plausible that sustained LE over time could potentially alter the brain networks engaged in speech understanding. Cross-sectional studies found that older adults with poorer hearing had reduced grey matter volume in auditory cortex compared with people with better hearing (Peelle 2018). Most notably, recent findings provided evidence that HI as well as daily life LE seems to be associated with grey matter loss in prefrontal brain regions (Rosemann and Thiel 2020). Further research is needed to elucidate the degree to which these neural changes occur and their relationship with cognitive difficulties.

1.7 THESIS OVERVIEW

1.7.1 Thesis aims and objectives

The literature review contained in this chapter suggests that listening may be an effortful task for CI users. However, due to the paucity of research, there is limited objective evidence of the LE experienced by CI users in everyday life. To date, questions such as “is listening through a CI inherently effortful?”, “how much LE do CI users experience in daily life compared with NH individuals?”, “how is the effort of listening through a CI best quantified objectively?”, and “what are the neural correlates of LE in the implanted population?” remain unanswered. Addressing these questions will shed light on the cognitive demands and challenges faced by CI users in everyday life. A better understanding of LE in the CI population could inform the development of objective assessments to quantify LE, as well as new interventions to make listening through a CI easier. Reducing LE is a key aspect of improving CI users’ quality of life and reducing any risk of social isolation and/or long-term cognitive decline (associated to HL).

To that end, the main goal of this thesis was to investigate the LE experienced by adult CI users in real-world listening scenarios, specifically, in those listening situations that involve social interactions either in person or remotely. Multimodal measures of LE were employed throughout the project, although particular attention was given to assessments that objectively quantify the cognitive demands of listening through a CI. Two novel approaches were proposed: the simultaneous acquisition of two physiological measures (fNIRS- based brain imaging and pupillometry); and a joint analysis of behavioural responses (accuracy and response time) capable of providing a “listening efficiency” outcome that reflects both intelligibility and LE. The suitability and practicality of these methods for the evaluation of effortful listening and performance in the implanted population was also examined.

The main objectives of this thesis were defined as follows:

i) To employ multimodal measures (cognitive, behavioural, subjective, and physiological) to assess the LE experienced by CI recipients (compared to NH controls) under realistic sound environments.

ii) To combine functional near-infrared spectroscopy (fNIRS)-based brain imaging with simultaneous pupillometry as a tool to explore the neural correlates of LE in CI users.

iii) To calculate a “listening efficiency” measure (based on decision-making models) that provides an objective evaluation of CI users’ listening performance, which reflects both the intelligibility achieved and effort expended.

iv) To examine correlations between participants’ listening efficiency and their cognitive scores, subjective ratings, and physiological responses.

v) To evaluate the feasibility and sensitivity of the objective assessments proposed (ii, iii) to characterise the mental exertion and listening performance of CI users.

vi) To identify which technological features make online communication (e.g., video calls) easier (i.e., less cognitively demanding) for CI users.

1.7.2 Thesis structure

Three studies were conducted throughout the duration of the project to investigate the LE that CI users may experience in everyday life when communicating in a busy café, through video calls, and in some scenarios that were prevalent during the COVID-19 pandemic.

The first study, reported in CHAPTER 2 and CHAPTER 3, was a laboratory experiment designed to assess LE in adult CI users (compared with age-matched NH controls) when they listen to speech under ecologically relevant levels of cafeteria background noise. To do so, behavioural measures (response time and accuracy), subjective ratings, pupillometry, and brain-imaging responses were collected simultaneously during the main speech in noise task. Additionally, individual differences in cognition and everyday listening experiences were examined by means of non-auditory cognitive tests and hearing questionnaires.

CHAPTER 2 presents and discusses the results of cognitive, subjective, and behavioural measures. Most notably, a novel analysis of behavioural responses using a hierarchical linear ballistic accumulator (LBA) model is proposed to quantify objectively participants' listening efficiency and its correlations with subjective and cognitive scores.

CHAPTER 3 reports results from physiological measures. Moreover, the physiological correlates of listening efficiency were also explored in this chapter, by assessing the correlation between participants' performance and their physiological reactions.

With the outbreak of the COVID-19 pandemic, new communication scenarios have emerged where restrictions such as wearing a facemask, pose additional listening challenges that greatly concern the HI community. Given that in-person research could not be conducted during much of the pandemic, we designed an online survey to explore the perceived listening difficulties of CI users under four commonly occurring communication scenarios, specifically, when listening to someone wearing a facemask, under social distancing guidelines, via telephone, and via video call. CHAPTER 4 presents the online survey and its results.

The survey results informed the development of an online test that further explored the cognitive demands of online communication from an objective perspective. CHAPTER 5 describes the online test whose aim was to investigate which technological features make online video calls easier (i.e., less cognitively demanding) for CI users. To accomplish it, the listening efficiency of CI users and age-matched NH controls was evaluated while completing a behavioural task of speech recognition under three video call presentation modes: when audio only, video (audio-visual), and video plus captions were available.

Finally, CHAPTER 6 provides an overview of the main findings and discusses the implications of this research, as well as the avenues for further exploration in this field.

CHAPTER 2. LISTENING EFFICIENCY: A NOVEL OUTCOME MEASURE THAT REFLECTS BOTH INTELLIGIBILITY AND EFFORT

Chapter adapted from: Perea Pérez F., et al (2023). Listening efficiency in adult cochlear-implant users compared with normally-hearing controls at ecologically relevant signal-to-noise ratios. Frontiers in Human Neuroscience. https://doi.org/10.3389/fnhum.2023.1214485

2.1 CHAPTER OVERVIEW

The assessment and understanding of LE have become a priority for the hearing research community in recent years. In real-life, listening is not only about hearing an attended target, but also about the amount of effort expended in doing so. Due to having to work with an impoverished auditory signal, CI users may experience reduced speech intelligibility and/or increased listening effort in typical real-world listening situations, compared to their NH peers. These two challenges to perception may be usefully integrated in a unique measure of listening efficiency: conceptually, the amount of accuracy achieved per unit of effort expended.

We describe a novel approach to quantifying “listening efficiency” based on the rate of evidence accumulation towards a correct response in a linear ballistic accumulator (LBA) model of choice decision-making. Estimation of this measure within a hierarchical Bayesian framework confers further benefits, including full quantification of uncertainty in parameter estimates, as well as improved estimation at the individual-subject level.

We applied this approach to examine, in a laboratory experiment, the speech-in-noise performance of a group of 24 cochlear implant (CI) users and a group of 25 approximately age-matched NH controls. Participants listened to reverberant target sentences in cafeteria noise at ecologically relevant signal-to-noise ratios (SNRs): +20 dB (“Easy”), +10 dB (“Medium”), and +4 dB (“Hard”). At a group level, the CI group was disproportionally affected by the background noise, showing much lower listening efficiency than the NH group, even in favourable acoustic conditions. At the individual level, within the CI group (but not the NH group), higher listening efficiency was associated with better cognition (i.e. working memory and linguistic-closure) and with more positive self-reported listening experiences, both in the laboratory and in daily life.

We argue that listening efficiency, measured using the approach described here, is: i) conceptually well-motivated, in that it is theoretically impervious to differences in how individuals approach the speed–accuracy trade-off (SATO) that is inherent to all perceptual decision making; and ii) of practical utility, in that it is sensitive to differences in task demand, and to differences between groups, even when speech intelligibility remains at or near ceiling level.

2.2 INTRODUCTION

CIs can partially restore hearing function to people with severe-to-profound HL. Their effectiveness has been widely proven in terms of improving outcomes such as speech recognition, overall quality of life (QOL), long-term well-being, and mental health (Buchman, Gifford, et al. 2020). Despite the benefits that CIs provide, not all CI recipients achieve the same level of performance in terms of speech intelligibility (Boisvert et al. 2020). Such variability is usually attributed to individual differences, including factors such as the duration of HL, age of implantation and the duration of CI use, among others. Regardless of any individual differences, the limitations of the CI technology impose additional challenges to speech perception, especially in noisy environments. The degradation of the auditory signal provided by implants can hinder the ability to distinguish and segregate sounds, making listening a highly taxing task (Başkent 2012; Winn et al. 2015). Indeed, the selective attention needed to stay focused on a desired speech target while ignoring irrelevant competing sounds, could lead CI users to experience LE (Strauss and Francis 2017; Wild et al. 2012). Certainly, previous research has found that CI users report experiencing high levels of LE and fatigue in everyday life (Alhanbali et al. 2017; Hughes et al. 2018; Rapport et al. 2020). This ongoing demand for increased mental exertion could have negative consequences for communication (Hetu et al. 1988), social participation

(Kramer et al. 2006; Mick et al. 2014; Nachtegaal et al. 2009; Pronk et al. 2011; Shukla et al. 2020), and long-term cognitive health (Lin et al. 2013; Pichora-Fuller, Mick, and Reed 2015).

In recent years, the assessment and understanding of LE has become a priority for the hearing science community (Francis and Love 2019; McGarrigle et al. 2014b; Pichora-Fuller et al. 2016). Different measures have been proposed to estimate the amount of effort exerted in a listening task. Attending to their nature, these measures are usually classified as physiological (brain activity and measures of the autonomic nervous system), subjective (self-reported and subjective assessments), cognitive (working memory and attention allocation) and behavioural (dual-task performance). Although these measures are sensitive to changes in participants' cognitive load and could provide a measurable index of the construct of LE, nowadays there is no standard method of measuring it (Alhanbali et al. 2019; Francis and Love 2019; McGarrigle et al. 2014b; Rudner et al. 2012). This is mainly because these measures are believed to provide complementary information that assess different underlying domains of the LE phenomenon (Alhanbali et al. 2019; Francis and Love 2019; Lau et al. 2019; Strand et al. 2018). For instance, self-report measures of LE are believed to reflect an affective or emotional response to how effortful listeners feel a task to be, whereas behavioural and cognitive measures are mostly considered measures of exerted effort, and thus are more dependent on changes in task demand. Physiological measures, such as pupil dilation, may capture changes in effort not only due to task demands but also due to participants' motivation and engagement, reflecting both the cognitive and affective dimensions of arousal (Francis and Love 2019; Pichora-Fuller et al. 2016).

Behavioural measures are perhaps the most commonly used assessments of LE due to their simplicity and feasibility, i.e., the tasks are easy to design, implement, and perform, and no special equipment is required. They provide an objective assessment of LE based on measurements of accuracy in task performance and speed of processing, often in the context of single or dual-task paradigms (Larsby et al. 2005; McGarrigle et al. 2014b). Most commonly, response time (RT, sometimes called reaction time) is measured as the fastest rate at which a cognitive task can be performed with reasonable accuracy (Phillips 2016). Behavioural assessments assume that both performance and speed of processing are reduced as the level of task difficulty increases. Previous research has considered behavioural measures to be effective and consistent to account for changes in LE (Deary 1994; Gatehouse and Gordon 1990; Hällgren et al. 2001; Houben et al. 2013; Larsby et al. 2005), suggesting them as appropriate evaluations of LE in clinical settings (Gosselin and Gagné 2010; Kaplan Neeman, Roziner, and Muchnik 2022). Nonetheless, measures of accuracy and RT need to be taken into account in combination since it is known that LE can still be experienced even when intelligibility performance is at or near ceiling (Houben et al. 2013; Pals et al. 2020; Winn and Teece 2022).

Considerable effort has been made in the field of experimental cognitive psychology to integrate both behavioural measures into a combined metric (Hughes et al. 2014; Liesefeld and Janczyk 2019; Vandierendonck 2017). In hearing sciences, this metric is usually interpreted as “listening efficiency” and considers both performance and response time in a listening task. Such integration is sought/preferred due to its consistency in terms of test-retest reliability which is greater than the ordinary analysis of response time and accuracy separately, which ignores any speed-accuracy trade-off (SATO) (Bakun Emesh et al. 2021; Salthouse and Madden 2008; Vandierendonck 2017). Some of the linear transformations proposed to combine both measures include metrics such as: the inverse efficiency scores (Townsend and Ashby 1983), the rate correct score (Woltz and Was 2006), the Linear Integrated Speed–Accuracy Score (Vandierendonck 2017), the bin score (Hughes et al. 2014), and the listening efficiency (Prodi and Visentin 2015; Prodi, Visentin, and Farnetani 2010; Visentin et al. 2017). However, these linear transformations do not consider the curvilinear relationship between speed and accuracy. Therefore, the estimations that they provide, although accurate in some cases, have some limitations that could lead to biased or noisy results when assessing both individual differences and group comparison (Bakun Emesh et al. 2021; Stafford et al. 2020).

To overcome this limitation, and given the need to have an integrated metric to better represent the underlying ability-trait while improving the retest reliability, in this chapter a RT decision model is proposed to perform a joint analysis of behavioural measures. Models of decision-making are able to characterise the SATO that is inherent to decision-making processes (Forstmann et al. 2011; Heitz 2014; Stafford et al. 2020). Such models not only provide a combined analysis of speed and accuracy data but also offer increased statistical power, making the most out of behavioural data (Stafford et al. 2020). Their use has become predominant in the cognitive psychology field and their effectiveness in characterising the underlying processes of rapid decision tasks has been widely demonstrated (Evans and Wagenmakers 2020; Forstmann et al. 2010, 2011; Forstmann, Ratcliff, and Wagenmakers 2016; Gomez, Ratcliff, and Perea 2007; Heathcote and Love 2012).

The main assumption behind these models is that decisions are made when enough evidence is accumulated in favour of a particular response option. Among all accumulator models available, we use here a hierarchical linear ballistic accumulator model (LBA: Brown & Heathcote, 2008). The LBA is a simplified version of these cognitive models and is classified as a race sequential sampling model in which the accumulation of evidence occurs linearly over time towards a common response threshold. They can be used to predict both response probabilities and response times in speeded decision-making paradigms. All possible choices (responses) are represented with independent evidence accumulators that gather evidence for each response. In this way, the decision made corresponds to the accumulator that first reaches the response threshold (Figure 2.1). The observed RT is assumed to be the sum of the decision and non-decision time. The decision time is the amount of time taken for the faster accumulator to reach the threshold, while the non-decision time *(t0)* is a constant value representing other non-decision processes.

The LBA model comprises different parameters that are related to different components of the underlying cognitive process that occur during the decision making (Evans and Wagenmakers 2020). These parameters shown in Figure 2.1 are: 1) the drift rate *(v)* which is the speed of evidence accumulation for each response option and is able to reflect both task difficulty and participants' efficiency in information processing; 2) the decision threshold *(b)* is the amount of evidence required to trigger a decision and is able to reflect task caution; 3) the starting point *(a)* is the amount of evidence that already exist for a particular response even before the accumulation of evidence starts, and represents any response bias that participants may have towards a particular response; 4) the non-decision time *(t0)* is the amount of time needed for other processes not related to the decision making, such as the speed of perceptual encoding (of a given stimulus) and response execution (e.g., button pressing); 5) the response caution, calculated as K+A/2, is the average amount of evidence required to reach a decision (response).

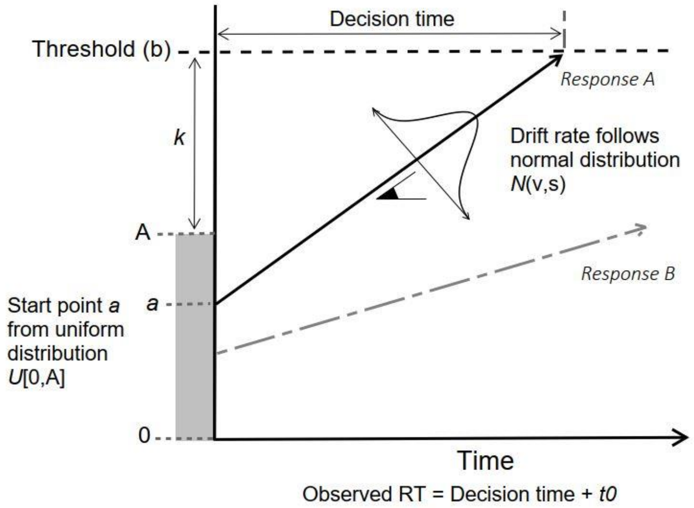


***Figure 2.1**. Graphical representation of the accumulation process assumed by LBA model, created based on Donkin et al. (2011) and Nishiguchi et al. (2019) studies. Racing LBA accumulators representing hypothetical responses A and B, where A is the selected response that first reached the evidence threshold (b).*

The LBA approach therefore allows dissociating RTs into the different processes that are involved in a response decision. Moreover, it uses hierarchical Bayesian statistics which provides clear advantages in estimating parameters at both individual and group levels (Katahira 2016; Liu, Lu, and Hoogenboom 2017; Nilsson, Rieskamp, and Wagenmakers 2011; Robert 2007b). Individual parameter estimates are constrained by group-level distributions, assuming that participants within each group are similar, but not identical to each other. Therefore, the model accounts for individual differences while identifying group commonalities. The probabilistic nature of the Bayesian approach also offers the ability to quantify the certainty of the parameters´ estimation (Annis and Palmeri 2018; Robert 2007a). It considers the entire response time distribution instead of single point estimates (e.g., mean, median). Likewise, results of model parameters are provided as full posterior probability distributions whose credible interval is computed by the 95% Highest Density Interval (HDI), which is the shortest interval that contains 95% of the mass posterior distribution (Hyndman 1996). In contrast to the orthodox confidence interval, one can be 95% confident that the true value of a particular parameter lies within the HDI interval (Lee and Wagenmakers 2014). Additionally, individual-level posterior distributions can be extracted to compute correlations between model parameters and other measures of interest. Following the plausible values approach, it is possible to obtain the sample plausible correlations that can then be generalised to the wider population (Ly et al. 2017; Ly, Marsman, and Wagenmakers 2018).

To take all the advantages of this approach, this study aimed to apply an LBA model to perform the analysis of the behavioural data collected in a laboratory experiment. The experiment was designed to assess the listening effort experienced by a group of adult CI users and a group of age-matched NH controls using a wide range of methods, including self-reported, cognitive, behavioural and physiological measures. Individual differences in cognition and everyday listening experiences were characterised by means of non-auditory working-memory and linguistic-closure tests, as well as standardised hearing questionnaires. In the main laboratory task, simultaneous behavioural, subjective, pupillometry, and brain-imaging measures were collected to assess the listening effort in an ecologically-relevant speech in noise task at three levels of difficulty (or signal-to-noise ratios).

Please note that the objective of this chapter is not to demonstrate the advantages of decision models over the analysis of RT and accuracy separately given that previous research has already addressed this (Bakun Emesh et al. 2021; Donkin et al. 2009; Stafford et al. 2020; White, Curl, and Sloane 2016). Instead, we exploit the LBA analysis to obtain a single metric that objectively characterises participants´ performance during listening tasks. We took as our primary performance metric the net drift rate since it provides an integrated estimation of the relationship between response time and accuracy (Donkin et al. 2009), and thus reflects both intelligibility and LE. We proposed this metric as a putative marker of participants´ listening efficiency and hypothesised that it would be sensitive to changes in listening performance across groups (between-subjects) and conditions (within-subjects). We expected that CI users would show inferior listening efficiency and greater self-reported effort than NH controls in both the laboratory speech-in-noise task and in questionnaires assessing daily life. Moreover, correlations between listening efficiency and participants' cognitive and subjective scores were explored to determine whether these could act as individual predictors of listening efficiency.

2.3 METHODS

The study was approved by the University of Nottingham Research Ethics Committee (reference: 247-1902).

2.3.1 Participant recruitment

Recruitment was carried out primarily through the Nottingham Biomedical Research Centre (Hearing Theme) Participant Database. The study was also advertised by national and regional hearing charities and organisations in the United Kingdom including the Royal National Institute for Deaf People (https://rnid.org.uk/) and the National Cochlear Implant Users Association (https://www.nciua.org.uk/).

A group of 24 CI recipients and a group of 25 age-matched NH controls volunteered to take part in the study (participant demographics in section 2.5.1). As per protocol, all participants were adults (aged 18 or over), right-handed as assessed using the Edinburgh Handedness Inventory (Oldfield 1971), English native speakers, with normal or corrected-to-normal vision (e.g., glasses), and no history of motor (e.g., cerebral palsy) or cognitive impairment (e.g., dementia or brain injury). Participants in the CI group were required to have at least 6 months of experience using their implant(s), and were tested in their best aided condition (e.g., with a HA in the contralateral ear if bimodal listeners). Participants in the NH group were confirmed to have normal hearing with a pure-tone audiometry air-conduction hearing screen (≤30 dB HL pure-tone average across 0.5, 1, 2, and 4 kHz). After providing informed consent, participants completed the experiment that lasted approximately two hours. Participants received an inconvenience allowance of £7.50 per hour and local travel expenses were covered (up to a maximum of £15).

2.3.2 Test procedure

After pure tone audiometry (PTA) and a short interview about participants' implantation experience (patient group only) were conducted, participants completed digitized versions of hearing questionnaires (SSQ12, EAS, FAS and HHQ) at their own pace on a touchscreen device. These questionnaires were administered to enquire about participants' daily life experiences (see section 2.3.3). Additionally, participants were asked to perform Reading Span (RSpan) and Text Reception Threshold (TRT) tests to characterise their working memory and linguistic abilities. Participants remained seated while performing the cognitive tests at approximately 45 cm from the external monitor where the sentences were displayed.

The main laboratory task was a hybrid block-event design divided into two runs where simultaneous behavioural, pupillometric, and optical brain-imaging measures were collected. Participants were instructed to listen to speech sentences masked by a continuous background noise and answer a simple yes/no question after each sentence, "was a given word presented visually on screen spoken in the sentence just heard?"

Each run lasted approximately 14 minutes (Figure 2.2) and consisted of three blocks of four minutes representing the three auditory experimental conditions. These listening conditions established the three levels of difficulty considered in the experiment and were defined by the SNR as "Easy" (20dB), "Medium" (10dB), and "Hard" (4dB). Each block contained 28 trials. The stimulus-onset asynchrony was randomly varied in the range of 6 to 10 seconds. Moreover, a constraint was defined so that no trial occurred within the first ten seconds of a block. In this way, participants were given some time to acclimatise to the background noise before the presentation of any trial.

There were two types of trials, sentence and null trials, each appearing 18 and 10 times per block, respectively, in random interleaved order. Each trial was comprised of a single sentence roughly 1.6 seconds in length, a post stimulus pause of 0.5 seconds and a yes/no decision task. In sentence trials, participants were instructed to listen to speech sentences masked by a continuous background noise. Then, they had to answer by pressing a button a simple yes/no question, whether a probe word (presented visually on screen) was featured in the sentence just heard.. Participants were encouraged to answer as accurately and quickly as possible and make their best guess when they were in doubt. Participants had three seconds to indicate their answer; otherwise, a missed response was recorded. The probability that the probe word had featured in the sentence was 50%, while in the remaining 50% of the trials a foil word was presented. Foil words were chosen to rhyme with the keyword, and, where possible, to be semantically plausible (e.g., in the sentence “The green tomatoes are small”, the keyword “green” might have been replaced with the foil word “clean”). No sentence was presented to the same participant more than once across the entire experiment. In null trials, the sentence was muted and only the noise was audible. In those cases, participants were instructed to submit a specific response either “press yes” or “press no”. Null trials acted as a noise baseline needed for the interpretation of brain imaging measures.

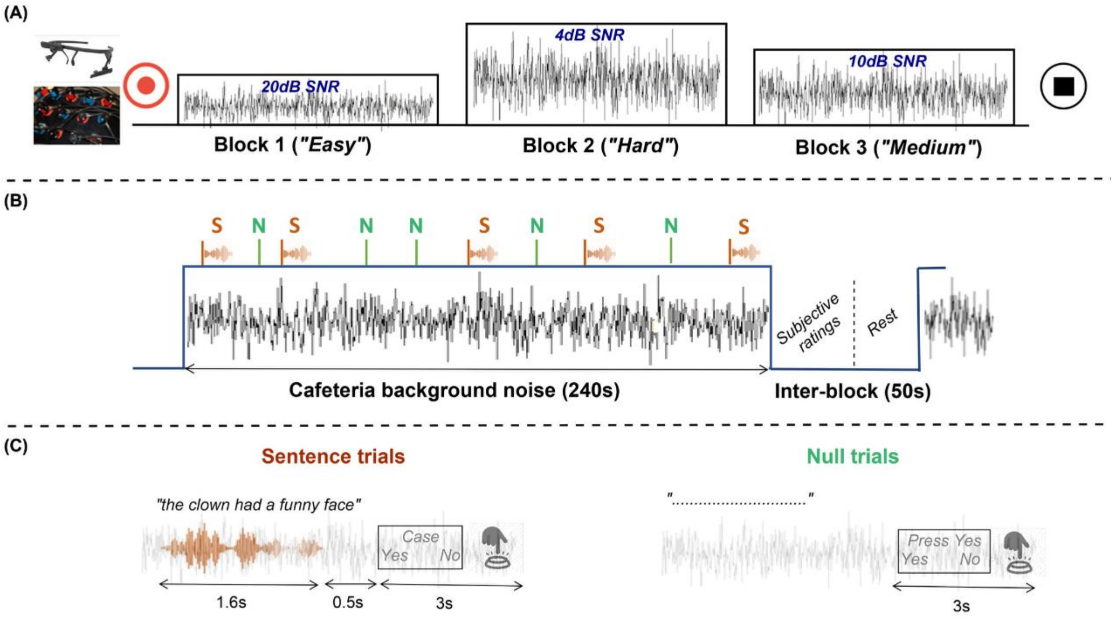


Figure 2.2. *Schematic representation of the speech-in-noise task. A) Example of an experimental run, whose blocks or experimental conditions are presented in easy, hard, and medium order (after randomisation). Physiological measures (fNIRS and pupillometry) are recorded for the duration of the entire run. B) Example of an experimental block, where sentence (S) and null trials (N) randomly presented, are masked by a continuous cafeteria background noise. During the inter-block pause, participants submit their subjective ratings. C) Example of sentence and null trials with their corresponding tasks: indicating whether a probe word was featured in the sentence or submitting a specific response as instructed in null trials. Both trial types had approximately the same duration.*

A silent baseline of approximately 50 seconds was also included between blocks. During this time and immediately after each block, participants were asked to report on their subjective listening experience during that block using visual analogue scales. Participants could respond anywhere along the 10 cm scale, with no intermediate marks or labels. The three questions that provided participants' task subjective scores were:

- Q1. Perceived effort: "How much effort was needed to understand the sentences?" (endpoints: "No effort" and "Extreme effort").

- Q2. Perceived intelligibility: "How many of the sentences did you understand?" (endpoints: "None of the sentences" and "All of the sentences").

- Q3. Task disengagement: "How often did you give up trying to understand the sentences?" (endpoints: "Never gave up" and "Always gave up").

Participants had 40 seconds to give their answers using a mouse as the input device. After this period, the questions disappeared, recording missed answers if no rating was reported. The total duration of the main task was approximately 30 minutes, however a break between runs was always offered for participant comfort. During testing, the researcher observed and took notes from a control room adjacent to the sound booth.

A short practice session was conducted before commencing data collection in which participants gained familiarity with the task and stimuli. During this practice, the researcher was present in the testing room to provide additional support when needed. The practice was designed to gradually instruct participants to perform the task. The practice session was conducted before the fNIRS and eye-tracker equipment were placed on the participant's head. All experimental programming was implemented in Matlab (MATLAB R2018b, The MathWorks Inc., Natick, MA, USA).

A momentary fatigue questionnaire (MFQ) was also completed by participants before and immediately after the main task to assess any change in participants' state of fatigue as a result of performing the task. Participants answered the question "How much fatigue (tiredness, weariness, problems thinking clearly) do you feel right now?" by putting a cross by hand anywhere on a numeric visual analogue scale divided in equal sections from 0 to 10, 0 being labelled as "None at all" and 10 as "Extreme fatigue".

2.3.3 Materials and Stimuli

The following tests and questionnaires were considered appropriate within the context of the experiment to characterise participants' cognitive abilities and listening experiences. These assessments were chosen due to their practicality; they are concise, intuitive, and easy to administer.

2.3.3.1 Hearing questionnaires

These questionnaires were delivered to assess participants' experiences and perceptions regarding hearing abilities, perceived handicap, LE, and fatigue in daily life.

- The effort assessment scale (EAS) is a questionnaire designed to measure self-reported listening effort in daily life of people with hearing loss (Alhanbali et al. 2017). The scale consists of six questions whose responses are provided on a visual analog scale from 0 indicating "No effort" to 10 "Lots of effort". Participants put a mark at any point of the scale that best represents their experiences. The ratings from all questions are added up to obtain the final EAS score, which is expressed in a range between 0 and 60, with higher scores indicating more effort.
- The fatigue assessment scale (FAS) (Michielsen et al. 2004) is a short and easy to administer scale that aims to measure fatigue in a general domain. It covers nine semantic categories: being bothered by fatigue, feeling physically tired, speed of getting tired, level of energy, concentration, inability to think clearly, quantity of daily activities, problems in starting things and feeling no desire to do anything. The FAS is formed of 10 short statements that are rated on a five-point likert scale divided into five answer categories: 1 = Never, 2 = Sometimes; 3 = Regularly; 4 = Often and 5 = Always. The scale score is calculated by summing up all items, taking into account that items 4 and 10 require reverse scoring. The total score of FAS ranges from 0 to 40, with higher scores indicating more fatigue. Although this questionnaire was not designed to assess specifically hearing related fatigue, other studies used it to evaluate the fatigue experienced by hearing impaired individuals (Alhanbali et al. 2017).
- The short version of the Speech, Spatial and Qualities of Hearing Scale (SSQ12) (Noble et al. 2013) was developed for use in clinical research and rehabilitation settings. It measures hearing ability across nine pragmatic domains: *Speech in Noise, Speech in Speech Contexts, Multiple Speech Stream*

Listening, Localization, Distance and Movement, Segregation, Identification of Sound, Quality and Naturalness, and Listening Effort. It consists of 12 questions with answers provided in a numeric visual analogue scale divided in equal sections from 0 to 10. Participants give their answers, by putting a mark at any point of the scale, considering that 0 means being unable and 10 being perfectly able to do or experience what is described in the question. A “non-applicable” box is also included for participants to indicate when a particular question is not relevant to their everyday experiences. Note that three items of the questionnaire address listening effort. The final score is calculated by averaging all the ratings reported and it ranges from 0 to 10, with higher scores indicating better hearing abilities.

- The Hearing Handicap Questionnaire (HHQ) was developed for the assessment of self-perceived hearing impairment related disability and handicap (Gatehouse and Noble 2004). Its psychometric properties are comparable with other questionnaires (the Hearing Handicap Inventory for Elderly (HHIE; Ventry and Weinstein, 1982) and the Hearing Handicap Inventory for Adults (HHIA; Newman et al., 1991). However, it provides additional advantages such as being able to test adults regardless of their age or listening capabilities (Thammaiah et al. 2017). The HHQ evaluates, in 12-items, the social restrictions and emotional distress caused by hearing impairment. Responses are scored using a five-point scale with equal intervals (almost, always, often, sometimes, rarely, never). All responses are averaged and scaled to provide a final handicap score that ranges from 0 to 100, with higher scores indicating greater handicap.

2.3.3.2 Cognitive tests

Two cognitive tests were used to assess individual differences in cognitive and linguistic abilities. Both tests were selected as they involve a non-auditory task that allows comparisons across participants and groups, regardless of their hearing status.

- The Reading Span Test (RSpan test) measures verbal working memory capacity with written stimuli (Daneman and Carpenter 1980). Among other working memory tests, it has been found to be more correlated to language comprehension, having been considered a good predictor of speech recognition performance in noise in hearing aid users (Akeroyd 2008; Rudner et al. 2008, 2009). Baddeley (1985)'s version was used in the study and consisted of the presentation of five-word sentences with a subject-verb-object syntax (e.g., "The captain" "sailed" "his boat"), half of which were semantically incorrect (e.g., "The train" "sang" "a song"). Participants were asked to read each sentence aloud and judge the semantic correctness of each sentence immediately after presentation. The sentences were grouped in three blocks of three, four, and five sentences respectively. After each block of sentences was presented, participants were asked to recall in the correct order either the first or the last words of every sentence in that block. A total of thirty-six sentences were presented in nine blocks. Prior to the test, one block of three sentences was presented in a practice session for participants to become familiar with the task. The Rspan score is the proportion of words that participants were able to recall correctly, and thus higher scores indicate greater working memory capacity.

- The Text Reception Threshold (TRT) test was developed by Adriana Zekveld and colleagues (Zekveld et al. 2007) as a visual analogue of the widely used Speech Reception Threshold. The test measures the "linguistic closure" ability to integrate and complete partially masked sentences. The test consists of reading aloud visually presented sentences that are partly masked by a bar pattern. The text that is not covered by the bar pattern represents the percentage of unmasked text that is modified throughout the test in an adaptive procedure, increasing or decreasing in 6% steps, according to participants' responses. For instance, when a sentence is correctly read the next sentence will show a wider bar pattern, decreasing the unmasked text by 6%; on the contrary, if the participant is not able to read the entire sentence, the next one will show a thinner bar pattern that reveals 6% more of unmasked text. No feedback is given during the test and participants are encouraged to make their best guess when in doubt. The TRT score is defined by the average percentage of unmasked text required to read 50% of the sentences correctly. Lower TRT scores indicates better performance.

Participants completed one practice session with 10 sentences and two TRT tests with 16 sentences each. For consistency, the sentences were presented in the same order to all participants and were obtained from three Bamford-Kowal-Bench (BKB) sentence lists, which were only used for the TRT test. Note that this experiment used a slightly modified version of the test, which was adapted and provided by Adriana Zekveld and colleagues. This version only scores key words within each sentence (three key words per sentence), in this way importance is given to words that provide more meaning within the context. For instance, if a non-key word (e.g., “the”) was not read correctly but all key words of the sentence were, then the overall sentence was scored as “correct”. The final rating was calculated for each participant as the averaged TRT score between the two tests performed.

2.3.3.3 Speech material and background noise

Speech material consisted of recordings of BKB sentences (Bench, Kowal, and Bamford 1979) spoken by a male talker. Seventeen sentence lists were available, each comprising 16 sentences. For each participant, a random subset of the available lists was selected and the sentences from those lists were randomly assigned to the three experimental conditions for each run. No participant was presented with the same BKB sentence more than once during the entire testing session. Prior to use, the sentences were convolved with the impulse response of the space in which the background noise was recorded. By doing so, the acoustic characteristics of the space, in particular the reverberation time (~1.4 s), were applied to the target speech sentences.

A real-world recording of a busy atrium café from the RealSpeechTM content library was used (with permission of Dr. Ian Wiggins and Dr. Mark Fletcher) as background noise. This “cafeteria” background noise was used to mask the target speech sentences. The difficulty of the listening task was manipulated by varying the level of speech relative to the level of the background noise (SNR), defining the three experimental conditions:

-Easy: +20 dB SNR. (Speech level 65dBA, noise level 45 dBA)
-Medium: +10 dB SNR. (Speech level 65dBA, noise level 55 dBA)
-Hard: +4 dB SNR. (Speech level 65dBA, noise level 61 dBA)

Each condition was presented in a separate block (i.e., the SNR was kept fixed throughout each four-minute block). The conditions were presented in random order for each run and participant. To avoid startling participants, the background noise was faded in and out gradually at the start and end of each block (fade duration 3 s).

It is worth mentioning that the type of background noise and SNRs assigned for each condition were chosen to be representative of everyday life sound scenarios. Particularly, the SNRs were selected based on previous studies that characterised the most commonly found real world SNRs of older adults with mild-to-moderate hearing loss (Smeds, Wolters, and Rung 2015; Wu et al. 2018). For instance, +4 dB SNR (the hard condition), was the average most common SNR found for “noisy” speech listening situations. Likewise, +10 dB SNR (the medium condition), was described to be the median SNR found in different listening environments, such as “home,” “indoors other than home,” and “outdoors”. Finally, +20dB SNR (the easy condition) was the most favourable condition found in everyday life scenarios and characterised as “very quiet situations”. Unlike previous studies, the SNRs used in the main task were not adjusted to each participant. This was intended to provide greater “realism” to the experiment, since people are not usually able to modify the noise levels encountered in real life scenarios.

2.3.4 Equipment

A touchscreen laptop connected to an external monitor was used to conduct the hearing questionnaires and cognitive tests. For this part of the experiment, participants were seated in the control room at approximately 45 cm from the external monitor, where the sentences were displayed.

The main behavioural task was conducted in a sound-attenuated room. Participants were seated at approximately 75 cm from a display screen with a loudspeaker (Model 8030A, Genelec, Iisalmi, Finland) mounted immediately above it. Auditory stimuli were presented in the free field. The sound pressure levels were measured at the listening position using a Brüel & Kjær Type 2250 sound level meter. Participants entered their responses using a mouse and a "RTbox" button box (Li et al. 2010)

Details about the settings and equipment used for pupillometry and brain imaging measures are provided in section 3.3.2 (CHAPTER 3).

2.4 STATISTICAL ANALYSIS

Participants' task subjective scores (perceived effort, intelligibility, and task disengagement) were calculated per each listening condition and averaged across runs using Matlab (MATLAB R2018b, The MathWorks Inc., Natick, MA, USA). Likewise, cognitive tests scores (RSpan and TRT tests) were calculated by custom applications implemented in Matlab and Delphi, respectively. These scores together with those resulting from the hearing and momentary fatigue questionnaires were normalized to 0-1 range before performing the statistical analysis in RStudio (Version 4.1.2; R Core Team, 2021). To facilitate comparison of hearing and cognitive test scores, reverse scoring was applied to the SSQ12 questionnaire and RSpan test so that greater scores represent worse hearing ability and less working memory capacity, respectively.

Group level differences were examined by computing Bayesian analyses using the R package brms (Bürkner 2017). The brms package implements Bayesian multilevel models using the probabilistic programming language Stan. The formula used to analyse hearing questionnaires, cognitive tests, momentary fatigue, and group age differences, regressed each outcome variable on the effect of Group (e.g., Age ~ Group), assuming unequal variances of both groups (e.g., sigma ~ Group). On the other hand, the formula for task subjective measures assessed the interaction "group per condition", taking into account participants' random effects by group (EF ~ 0 + Intercept + Group:Condition + (1 | gr(Participant, by = Group))). Ordered beta regression (Kubinec 2022) was set as the custom family distribution to model participants' responses to hearing questionnaire, cognitive tests and task subjective measures. Such distribution was explicitly designed for survey data where slider and visual analog scales (with both lower and upper bounds) are used. The Gaussian family was chosen to perform linear regression for outcome variables such as age and momentary fatigue measures. The latter was analysed on the difference post-vs pre experiment momentary fatigue scores. Prior distributions for ordered beta regression models were set as defined by Kubinec's study (Kubinec 2022), whereas default flat priors were used on the effect of age and momentary fatigue scores. Posterior distributions were estimated using the Markov Chain Monte Carlo (MCMC)(Van Ravenzwaaij, Cassey, and Brown 2018) algorithms, whose convergence was measured by the potential scale reduction factor R-hat(R) (Brooks and Gelman 1998) over four separate chains, each with 2000 warmup iterations followed by another 2000 post-warmup iterations. Posterior predictive checks were performed to ensure that the models' predictions adequately fit the data.

The model conditional effects, predicted means and 95% credible intervals (CrIs), were reported per group and condition (when relevant). Effect sizes were also calculated, when relevant, using Cliff's Delta statistics on the posterior distribution of the model's predicted data. This non-parametric effect size measure was chosen for its suitability to analyse ordinal data (Likert scales), which reduces the influence of outliers or groups' variance differences.

A hierarchical Linear Ballistic Accumulator model (LBA) was performed to analyse participants' behavioural responses following Gunawan's approach (Gunawan et al. 2020), that allows modelling intercorrelated individual level LBA parameters using a multivariate normal prior distribution. The LBA model was implemented in Stan as described in Annis et al.'s article (Annis, Miller, and Palmeri 2017), using a non-centred parameterisation to efficiently explore the posterior parameters' distributions. Missing response times (late responses>3s) were treated as parameters estimated within the model, assuming that their response accuracy would have been at chance level. The model scaling constraint was set so that the between-trial variability in the drift rates was fixed to one (sv=1). The model considered fifteen free parameters, two evidence accumulators, one per each response option (correct vs incorrect), that were allowed to vary across SNR conditions (Easy, Medium, Hard) and trial type (Sentence vs Null). The remaining parameters A, b, and t0 were allowed to vary across groups but were fixed across trials. Uninformative priors (normal and Student-t distributions) were used for the model parameters. Posterior distributions were estimated using MCMC algorithm, over four separate chains, each with 1000 warmup iterations followed by another 2000 sample iterations (8000 draws in total). Posterior predictive checks were used to assess the agreement between model predictions and observed data. Effects were assessed using 95% Highest posterior density intervals (HDI; HDInterval package) that acted as the 95% credible intervals (CrI) of the average posterior parameters.

Differential drift rates were calculated per each group as the difference between correct and incorrect responses. This parameter indicates the rate at which a listener preferentially accumulates evidence towards a correct response and was considered a putative indicator of participants' listening efficiency. Thus, faster accumulation of evidence towards correct answers was interpreted as greater listening efficiency, that is, greater ability for correctly recognising speech. Differential drift rates were calculated for both trial types to examine group differences during the listening (sentence trials) and task execution (null trials) processes.

Moreover, individual-level differential drift rates in sentence trials were extracted (per participant) and averaged across conditions. A correlation analysis following the plausible values approach (Ly et al. 2017) explored relationships between participants' listening efficiency and task subjective, cognitive, and hearing questionnaires scores. Correlations between 8,000 posterior draws per individual's drift rates and their scores on these measures were computed, resulting in a distribution of plausible correlations. The posterior distribution of the population correlation was also calculated (Ly et al. 2018), and each group's correlation mean and 95% credible interval were reported. Attending to the Pearson correlation coefficient, the magnitude of sample correlations was interpreted as either weak (r≤0.39), moderate (0.40≤r≤0.59), or strong (r≥0.6). Plausible population correlations (plausible p) were considered reliable when the 95% credible interval did not contain zero or when a high degree of certainty (>90%) suggested that the true value (plausible p) was different from zero.

2.5 RESULTS

2.5.1 Participant demographics and hearing profile

Both groups of participants had a similar age range; CI users: 20-84 years and NH controls: 20-79 years. The age similarity of both groups was confirmed by the results of the analysis (Figure 2.3) that showed similar predicted means and overlapping 95% CrIs (CI: 60.5 [53.8,67.1], NH: 55.9 [50, 62]) between groups. In terms of gender, both groups were similarly distributed with 42% and 44% of males in the CI and the NH group, respectively.

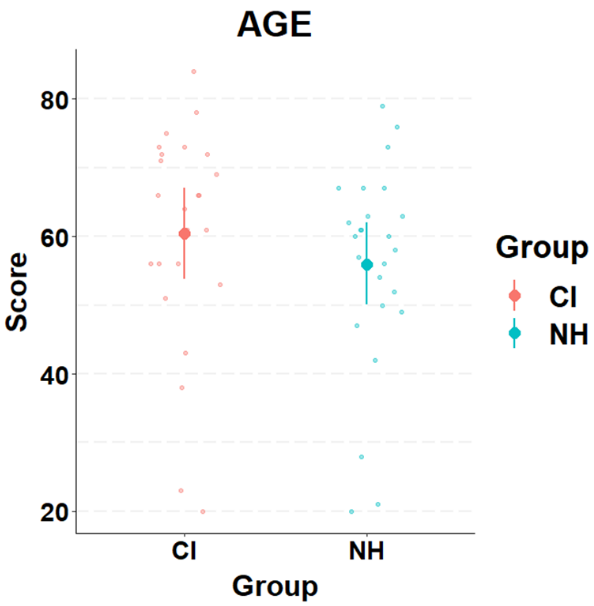


Figure 2.3. *Model conditional effects over raw data for participants' age by group (Age~Group). The error bars display 95% credible intervals; the bold dots represent posterior means, and the small dots represent the raw data.*

Participants in the control group had normal (or near-normal) hearing as assessed using air-conduction PTA screen across frequencies 0.5, 1, 2, and 4 kHz in both ears (22 adults with an average threshold ≤20 dB HL and three adults with an average threshold ≤30 dB HL). Participants in the CI group had an average of 8 years of CI experience (range 1-24 years). According to hearing device configuration, there were 11 bimodal listeners, 11 unilateral and 2 bilateral CI recipients. Most participants had severe-to-profound HL in the non-implanted ear as revealed by mean audiometric thresholds (in dB HL) (Unilateral: 93,100,100,100; Bimodal: 83, 82, 85, 92). Likewise, low-frequency residual hearing in the implanted ear(s) was hardly preserved [Unilateral (M: 100, SD: 1.5); Bimodal (M: 95, SD: 12.3); Bilateral (M: 99, SD: 1.8) dB HL]. Twenty CI users reported having developed HL after language acquisition while four were pre-lingually deafened. In all cases, CI participants were able to perform all listening tests included in the study as well as maintain conversations and communicate effectively with the researcher. Demographic information for CI participants is listed in Table 2.1.

Table 2.1. *Demographics of CI participants.*

CI Participant	Gender	Age (years)	CI Manufacturer	Hearing devices (implanted side)	Aetiology of Deafness	Years CI Exp.
1	M	66	Cochlear Nucleus 6	Unilateral (R)	Virus or disease	5
2	M	66	Cochlear Nucleus 6	Unilateral (R)	Born deaf/ Not Known	15
3	M	56	Cochlear Nucleus 7	Bimodal (R)	Not known	6
4	F	61	Cochlear Nucleus 7	Unilateral (L)	Nerve damage	5
5	F	61	Cochlear Nucleus 7	Bimodal(R)	Not known	6
6	M	38	AB	Bimodal(L)	Born deaf/ Not Known	8
7	F	56	Cochlear Nucleus 7	Unilateral (R)	Not known	4
8	F	69	Cochlear Nucleus 7	Unilateral (L)	Not known	1
9	M	23	Cochlear Nucleus 7	Unilateral (R)	Ototoxicity	21
10	F	20	Med-EL	Bilateral CIs	Born deaf/ Not Known	8
11	M	56	Cochlear Nucleus 6	Bimodal (R)	Born deaf/ Not Known	4
12	M	84	Cochlear Nucleus 6	Bimodal (L)	Virus or disease	3
13	F	75	AB	Bimodal (L)	Virus or disease	13
14	M	66	Med-EL	Unilateral (R)	Virus or disease	8
15	F	64	Med-EL	Bimodal (R)	Genetics	9
16	F	78	Med-EL	Unilateral (R)	Virus or disease	8
17	F	73	Med-EL	Bilateral CIs	Born deaf/ Not Known	23
18	F	72	Cochlear Nucleus 7	Unilateral (R)	Virus or disease	2
19	M	73	Cochlear Nucleus 6	Bimodal (L)	Genetics	4
20	F	53	Cochlear Nucleus 6	Bimodal (L)	Virus or disease	22
21	M	51	Cochlear Nucleus 6	Bimodal (L)	Ototoxicity	5
22	F	72	Cochlear Nucleus 7	Unilateral (L)	Not known	24
23	F	71	Cochlear Nucleus 7	Unilateral (R)	Not known	5
24	F	43	Cochlear Nucleus 7	Bimodal (L)	Not known	2

Table abbreviations per variable. Gender: F for female and M for male. CI manufacturer: AB for Advanced Bionics. Implanted side: R for right and L for left. The variable "Years CI Exp." refers to years of CI usage experience.

2.5.2 Cognitive and hearing function

Participants in both groups achieved similar scores on cognitive tests (Figure 2.4). Although the difference in scores between groups was slightly greater in the RSpan test compared to the TRT test, the model 95% CrIs show considerable overlap between groups. The model predicted means and 95% CrIs for the the RSpan test were 60.3 [54.9, 65.1] and 55.7 [50.6, 60.6], for the CI and NH groups respectively, whereas for the TRT they were 55.9 [52.3, 59.2] and 55.1 [51.6, 58.4].

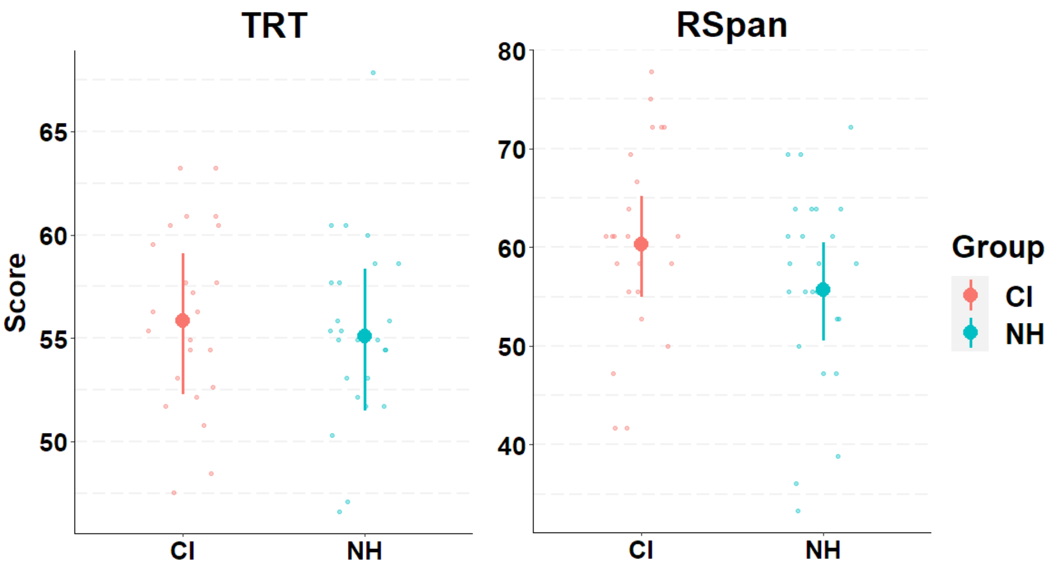


Figure 2.4. *Model conditional effects over raw data for participants' cognitive tests results by group (e.g., TRT~ Group). Abbreviations refer to Text Reception Threshold (TRT) and Reading Span (RSpan) tests for both groups of cochlear implant (CI) and normally hearing (NH) participants. The error bars display 95% credible intervals; the bold dots represent posterior means, and the small dots represent the raw data. Scores for TRT and RSpan tests range between 0-100, with greater scores indicating worse performance. Note that RSpan scores were reversed so that greater scores represent less working memory capacity.*

The results of hearing questionnaires however showed greater differences between groups, with CI users reporting greater hearing difficulties on all questionnaires compared to their NH peers (Figure 2.5). Strong effects of group were evident in the EAS, HHQ, and SSQ12 as revealed by non-overlapping 95% CrIs. Indeed, CI users' EAS scores were double those reported by NH controls, suggesting that participants in the CI group were greatly affected by LE in daily life. Likewise, a difference of 3.3 points (out of 10) in the SSQ12 scores between groups suggested that CI users exhibited worse hearing abilities compared to their NH peers. HHQ scores showed a difference of 35 points (on a 100-point scale) between groups, which again is interpreted as higher self-perception of hearing disability and handicap of CI users compared to NH controls. The model predicted means and 95% CrIs for the CI group in these questionnaires were: EAS (42.1, CrI:[36, 47.3]); FAS (8.7, CrI:[6.3, 11.6]); HHQ (43.4, CrI:[33.3, 53.7]); and SSQ12 (5.2, CrI:[4.5, 6]). The NH group predicted results in these questionnaires were: EAS (19.9, CrI:[14.3, 26.1]); FAS (7, CrI:[5, 9.4]); HHQ (8.8, [4.3, 15.4]); and SSQ12 (1.9, CrI:[1.4, 2.4]).

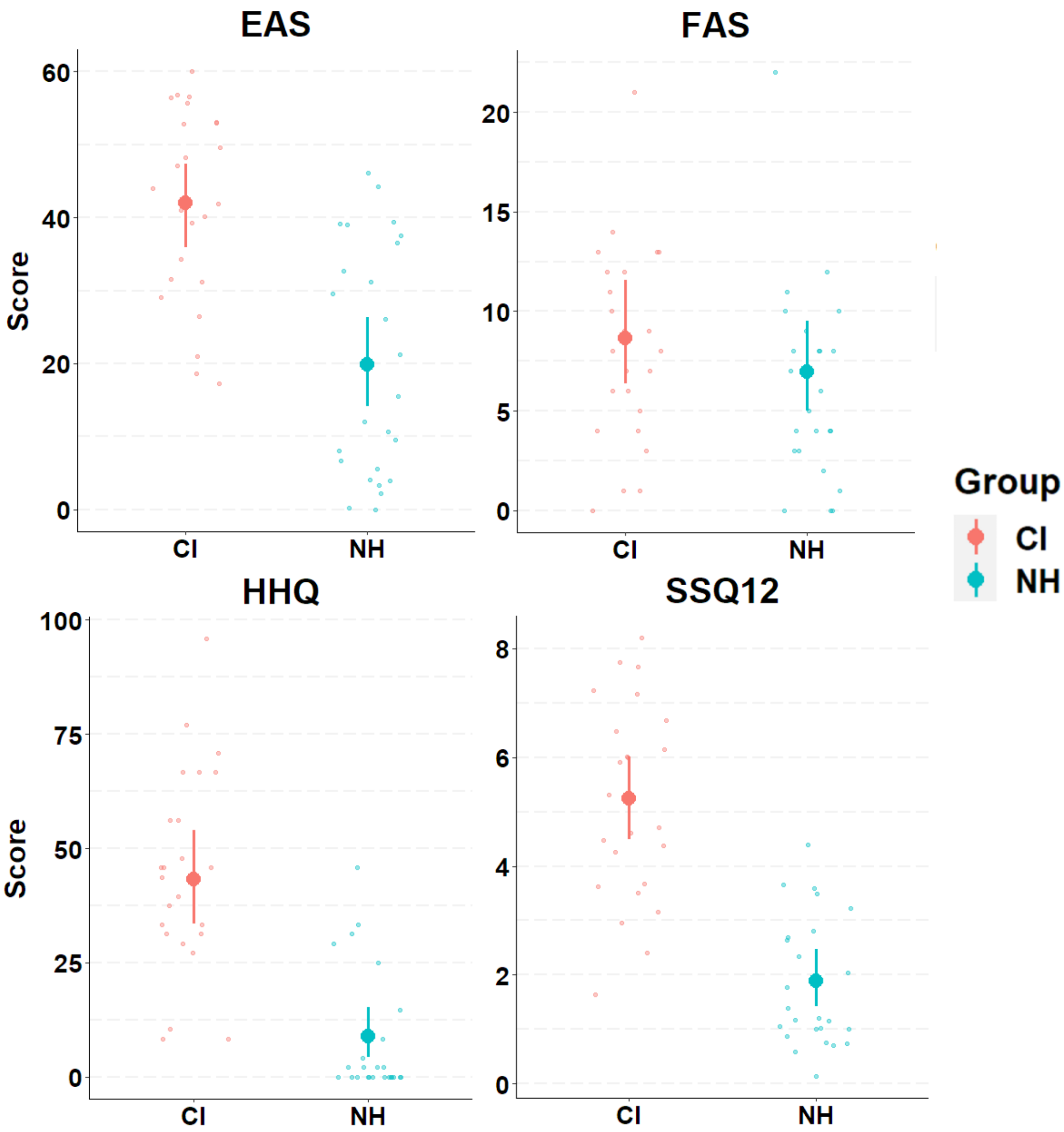


Figure 2.5. *Model conditional effects over raw data for participants' hearing questionnaires results by group (e.g., EAS ~ Group). The error bars display 95% credible intervals; the bold dots represent posterior means, and the small dots represent the raw data. The effort assessment scale (EAS) questionnaire has a score range between 0-60 points. The fatigue assessment scale (FAS) ranges between 0-40 points. The hearing handicap questionnaire (HHQ) has a score range of 0-100. The short version of the speech, spatial and qualities of hearing scale (SSQ12) was reverse scored, with a total range of 0-10 points. Greater scores in all questionnaires indicate greater hearing difficulty.*

Overall, low scores of fatigue were reported by participants before [CI (M: 2.7, SD: 1.9); NH (M: 1.5, SD: 1.8)] and after [CI (M: 4.7, SD: 2.8); NH (M: 2.8, SD: 2.2)] performing the main laboratory task. The change in participants' state of fatigue due to the experiment (Figure 2.6) was similar in both groups as revealed by the model predicted means and 95% CrIs (CI: 2 [1.2, 2.8], NH: 1.3 [0.8, 1.8]).

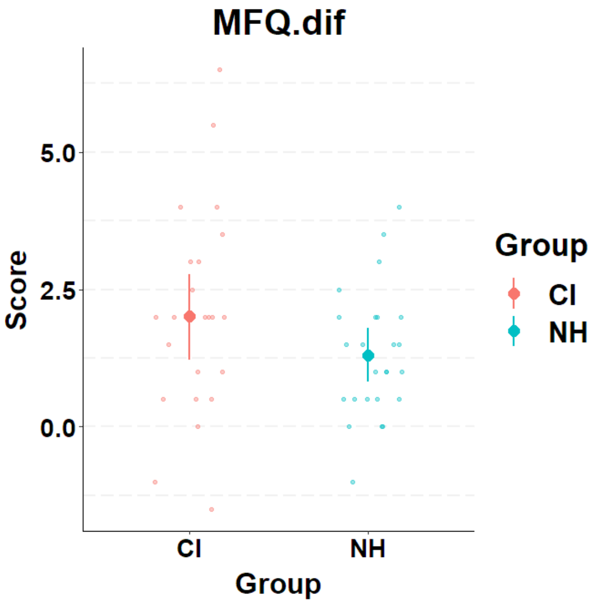


Figure 2.6. *Model conditional effects over raw data for the difference post-vs-pre experiment in participants' momentary fatigue questionnaire (MFQ) by group. The error bars display 95% credible intervals; the bold dots represent posterior means, and the small dots represent the raw data. Scores of MFQ were measured on a 10 points scale.*

Cliff's Delta effect size calculation (Figure 2.7) confirmed large group effects in EAS (d=0.7), HHQ (d=0.9), and SSQ12 (d=0.7) questionnaires, with group-difference posterior distributions that do not contain zero. This suggests that CI users experienced significantly greater hearing difficulty in daily life compared to NH controls. Cognitive measures, however, showed a weak effect of group on both the TRT (d=0.08) and RSpan (d=0.2) tests, suggesting that on average participants in both groups exhibited similar working memory capacity and linguistic closure abilities. Likewise, the effect of group in fatigue scores was small in both the FAS (d=0.2) and the MFQ post-vs-pre-experiment (d=0.25). These weak effects were also associated with a large uncertainty as expressed by the width of the credible intervals.

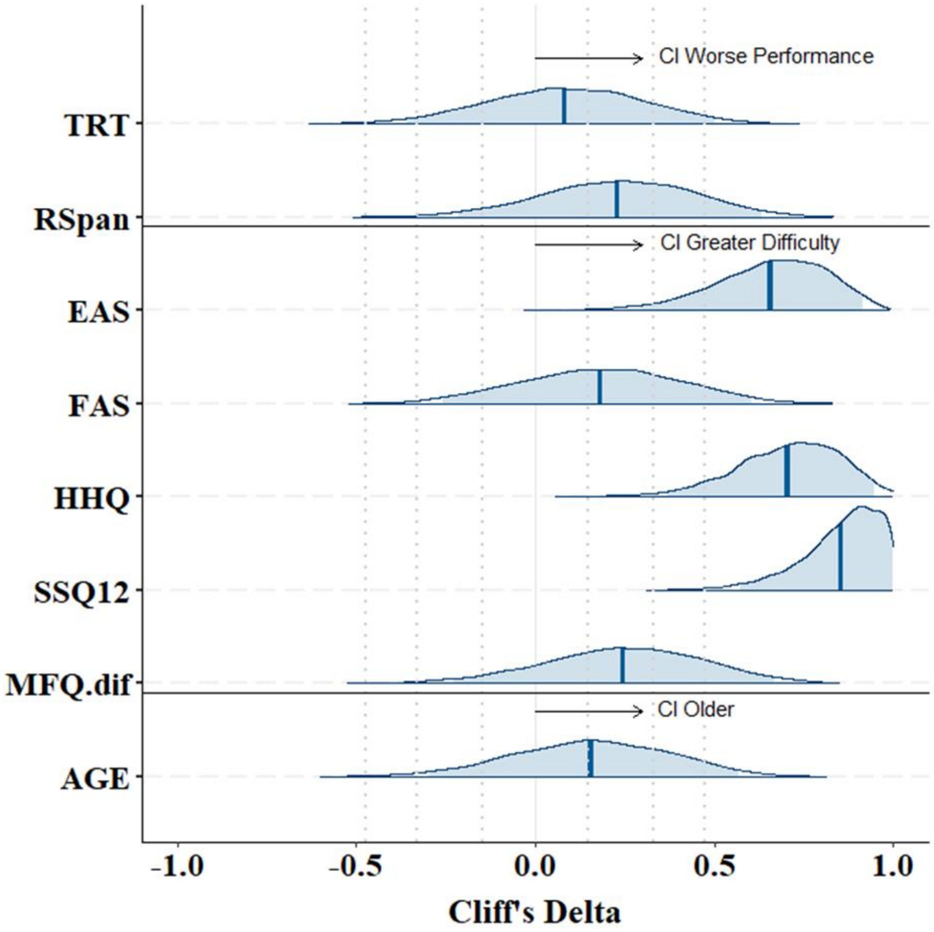


Figure 2.7. *Group-difference Cliff's Delta effect sizes with 95% credible intervals on the posterior distributions of cognitive tests (TRT, RSpan), hearing questionnaires (EAS, FAS, HHQ, SSQ12, MFQ.dif), and participants' age. Positive Cliff's Delta values indicate greater scores/results of participants in the CI group compared to the NH group (CI scores>NH scores). Abbreviations refer to Text Reception Threshold (TRT), Reading Span (RSpan) tests, Effort Assessment Scale (EAS), Fatigue Assessment Scale (FAS), Hearing Handicap Questionnaire (HHQ), short version of the Speech, Spatial and Qualities of hearing scale (SSQ12), and Momentary Fatigue Questionnaire (MFQ.dif) on the difference post-vs-pre experiment.*

2.5.3 Task subjective results

2.5.3.1 Self-reported listening effort

Participants in the CI group reported greater levels of perceived listening effort during the behavioural task compared to those in the NH group in all SNR conditions. The model predicted means (Easy [CI M: 0.48, NH M: 0.02], Medium [CI M: 0.69, NH M: 0.04], Hard [CI M: 0.82, NH M: 0.1]) and 95% CrIs on the interaction "Group x Condition" confirmed this as shown in Figure 2.8. Such a strong effect of group was evident as revealed by 95% CrIs that did not overlap in any condition. These results suggest that CI users on average perceived at least 50% more LE than NH controls at all task difficulty levels. Although more dramatic in the patient group, the increasing trend in participants' perceived effort as SNR worsened was present in both groups.

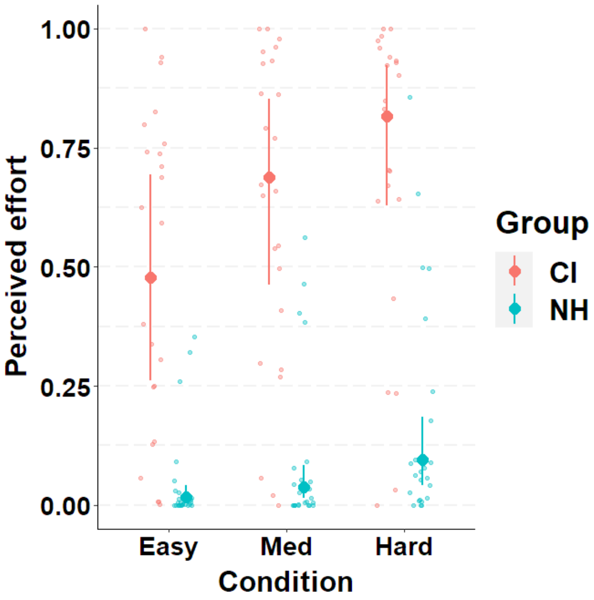


Figure 2.8. *Model conditional effects over raw data for participants' self-perceived listening effort by Group and Condition (EF ~ 0 + Intercept + Group:Condition + (1 | gr(Participant, by = Group))). The error bars display 95% credible intervals; the bold dots represent posterior means, and the small dots represent the raw data. Scores were measured in a 0-1 scale.*

2.5.3.2 Self-reported intelligibility

The analysis of task perceived intelligibility also indicated considerable differences between both groups of participants (Figure 2.9). A ceiling effect was observed in the NH group in all experimental conditions as shown by the model predicted means (Easy M:0.99, Med M:0.99, Hard M:0.98), with extremely narrow 95% CrIs (± 0.01). The patient group (CI), however, reported significantly inferior levels of intelligibility in all conditions (Easy [M: 0.83, CrI: (0.69, 0.92)], Med [M:0.74, CrI: (0.655, 0.86)], Hard [M:0.59, CrI: (0.39, 0.76)]), with such difference being more dramatic as the level of task difficulty increased.

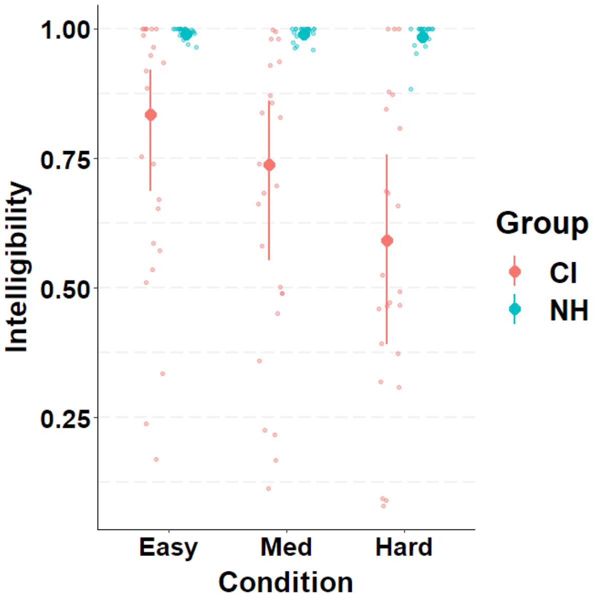


Figure 2.9*. Model conditional effects over raw data for participants' self-perceived intelligibility by Group and Condition (IN ~ 0 + Intercept + Group:Condition + (1|gr(Participant, by = Group))). The error bars display 95% credible intervals; the bold dots represent posterior means, and the small dots represent the raw data. Scores were measured in a 0-1 scale.*

2.5.3.3 Self-reported task disengagement

Overall, low levels of task disengagement were reported by all participants in all experimental conditions. This floor effect in task disengagement was observed in the model predicted means (Easy [CI group M: 0.017, NH group M: 0.006], Med [CI group M: 0.02, NH group M: 0.005], Hard [CI group M: 0.04, NH group M: 0.004]) and narrow 95% CrIs (Figure 2.10). A small effect of group was only present in the hard condition where the 95% CrIs did not overlap.

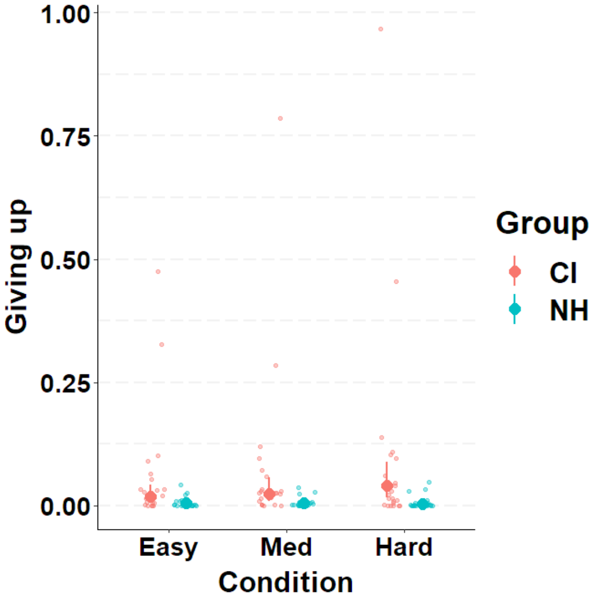


Figure 2.10. *Model conditional effects over raw data for participants' self-perceived task disengagement by Group and Condition (TD ~ 0 + Intercept + Group:Condition + (1|gr(Participant, by = Group))). The error bars display 95% credible intervals; the bold dots represent posterior means, and the small dots represent the raw data. Scores were measured in a 0-1 scale.*

2.5.4 Behavioural results

The LBA model was fed with 8064 observations, corresponding to 47 participants who completed 168 trials each (two runs x three conditions x 28 trials [18 sentence + 10 null trials]), and two participants who only completed one run each (84 trials). Sixty-three observations were treated as missing responses within the model. Missing responses occurred when participants took more than 3 seconds to submit their answer and therefore, a not-known response was assumed. Satisfactory convergence was found for all estimated parameters as revealed by the full traces' plots and Rhats' range [0.99, 1.01]. Posterior predictive checks showed an adequate fit of the model´s predictive RT distributions to the observed data (Figure 2.11). The mean difference between predicted and observed data was 0.1% and 0.5% for the NH and the CI group, respectively, across conditions and trial types.

As can be seen in Figure 2.11, very few errors (plotted as negative) were made by participants during the behavioural task (on average 0.5% incorrect responses were submitted by NH participants and 9% by CI users across conditions and trial types).

The high levels of accuracy were not surprising considering that the behavioural task was a yes-no task in which correctness of responses are always at least at a chance level (50%). Moreover, the listening conditions, all with positive SNRs, were considerably easy for NH listeners and more challenging for CI recipients, hence the difference in error rate between groups. To take into account this difference, the results are expressed using differential drift rates (vDiff= vcorrect-vincorrect).

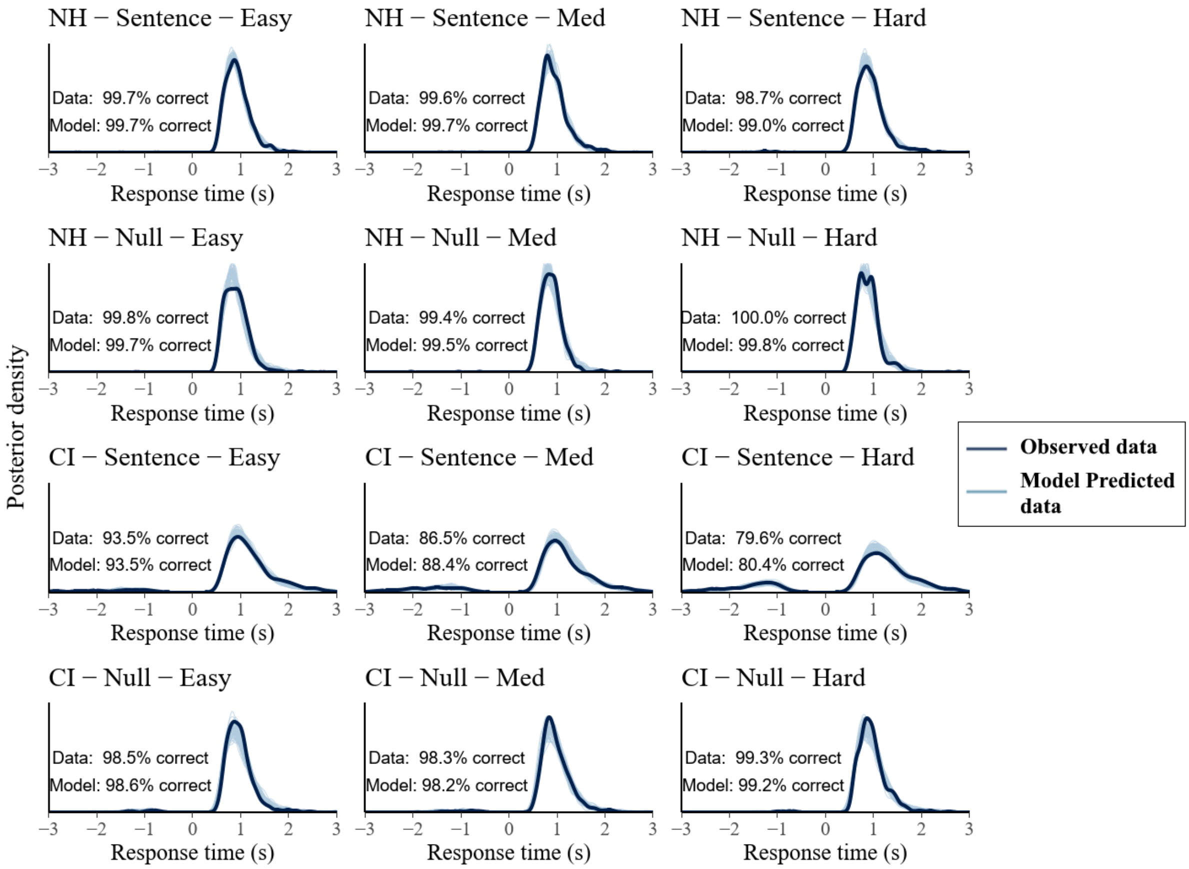


Figure 2.11. *Posterior predictive checks per group (NH and CI group shown at the two top and bottom lines, respectively), trial type (sentence vs null trials), and condition (Easy, Med, and Hard conditions shown from left to right columns). Participants' response time (RT) for incorrect responses are plotted as negative. Solid dark lines represent the observed data and light blue lines represent the model predicted data (8000 draws).*

Significant group differences were found in the posterior distributions of the LBA´s drift rates parameters in sentence trials (Figure 2.12). Overall, NH participants showed faster accumulation of evidence towards correct answers in sentence trials in all experimental conditions compared to CI users. Such a strong effect of group was confirmed by posterior distributions that do not overlap in any SNR condition. This can also be seen in the between-group difference column of Table 2.2 with means and 95% CrIs that do not contain zero. Conversely, no effect of group was found in null trials. Although CI users´ performance in null trials was slightly inferior compared to NH controls, such difference did not reach significance as revealed by the between-group difference in null trials’ drift rates (Table 2.2). It comes as no surprise that differences in performance between groups were only present in sentence trials, since null trial tasks did not require active listening but just following instructions instead.

The increase in task difficulty was also noticeable within participants´ performance across conditions in sentence trials. The speed of evidence accumulation was reduced as the SNR became less favourable (Figure 2.12). This tendency was present in both groups and confirmed by negative slopes of sentence trials’ drift rates across conditions (see ‘vDiff.Sentence.Slope’ in Figure 2.13 and Table 2.2). The effect was stronger in the CI group as revealed by steeper negative slopes across conditions (CI M: -0.61; NH M:-0.26). These results suggest that CI users may be affected to a greater extent by the increase in task difficulty than their NH peers. As expected, participants’ performance in null trials was not significantly affected by the increase in task difficulty. Indeed, their performance in null trials was similar across conditions as revealed by posterior distributions of slopes that contained the “zero slope” value in both groups.

Parameters such as the non-decision time (t0) and response caution (K+A/2) were almost identical in both groups (Figure 2.12 and Table 2.2). It is assumed therefore that participants in both groups had similar non-decision timings and levels of caution. In other words, they spent the same amount of time in non-decision processes and needed the same amount of evidence to make a choice.

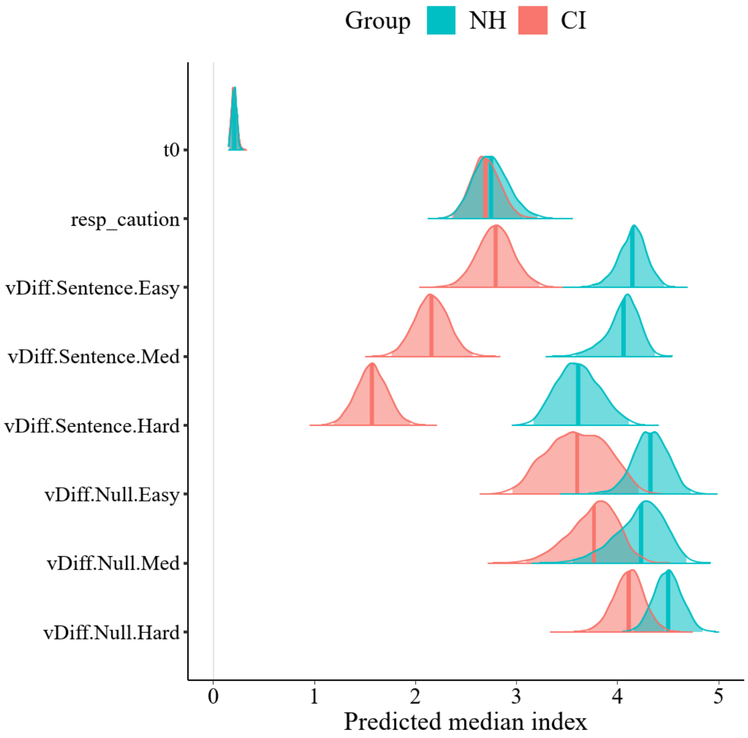


Figure 2.12. *Posterior group comparison in LBA model's parameters: t0, response caution, and differential drift rates (vDiff) per trial type (Sentence,Null), and condition (Easy, Med and Hard). Solid lines in posterior distributions represent the predicted median index for each parameter.*

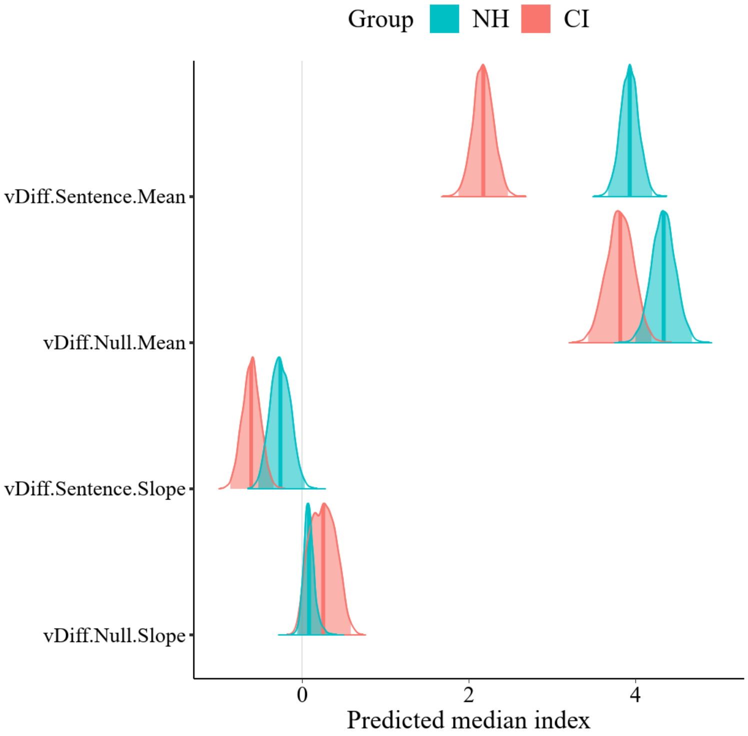


Figure 2.13. *Mean value and Slope across conditions of LBA model's differential drift rates per trial type (Sentence, Null). Solid lines in posterior distributions represent the predicted median index for each parameter.*

***Table 2.2**. Means and Highest Density Intervals (HDI) of the LBA model's parameters posterior distributions by group. The effect of group can be seen in the between-group difference columns (last three columns). The HDI is interpreted as the 95% credible intervals (CrI) of posterior distributions.*

	NH Group			**CI Group**			**Between-Group Difference *(NH-CI)***		
	Mean (v)	**95%CrI Lower**	**95%CrI Upper**	**Mean (v)**	**95%CrI Lower**	**95%CrI Upper**	**Mean (v)**	**95%CrI Lower**	**95%CrI Upper**
t0	0.21	0.15	0.25	0.21	0.15	0.26	0.00	-0.08	0.07
resp_caution	2.76	2.38	3.21	2.70	2.33	3.03	0.06	-0.45	0.57
vDiff.Sentence.Easy	4.14	3.75	4.46	2.79	2.35	3.23	1.35	0.79	1.88
vDiff.Sentence.Med	4.04	3.61	4.39	2.16	1.75	2.55	1.88	1.31	2.43
vDiff.Sentence.Hard	3.62	3.18	4.12	1.57	1.20	1.95	2.05	1.45	2.65
vDiff.Null.Easy	4.32	3.86	4.74	3.60	2.97	4.22	0.72	-0.01	1.48
vDiff.Null.Med	4.20	3.59	4.70	3.74	3.13	4.26	0.46	-0.28	1.17
vDiff.Null.Hard	4.50	4.16	4.83	4.10	3.72	4.47	0.39	-0.08	0.90
vDiff.Sentence.Mean	3.93	3.67	4.20	2.17	1.89	2.48	1.76	1.35	2.14
vDiff.Null.Mean	4.34	3.99	4.67	3.81	3.42	4.18	0.52	0.05	1.02
vDiff.Sentence.Slope	-0.26	-0.54	0.02	-0.61	-0.86	-0.35	0.35	0.01	0.75
vDiff.Null.Slope	0.09	-0.08	0.29	0.25	-0.06	0.57	-0.16	-0.52	0.18

Table parameters are t0, response caution (resp_caution), and differential drift rates (vDiff). Drift rates are described in the format "vDiff. trial type. Condition" and "vDiff. trial type. Mean or Slope across conditions".

2.5.5 Correlation analysis

Relationships between participants' listening efficiency and their scores on cognitive, subjective, and hearing questionnaires were explored following a plausible values approach. To do so, differential drift rates in sentence trials were extracted per participant and averaged across conditions (8000 draws each). Then, the relative importance of predictor variables (age, hearing questionnaires and cognitive scores) were calculated in a multiple linear regression model using the Relaimpo R package (Grömping 2006). Figure 2.14 shows the relative contribution of these predictor variables to individual-level listening efficiency (sentence differential drift rates) by group.

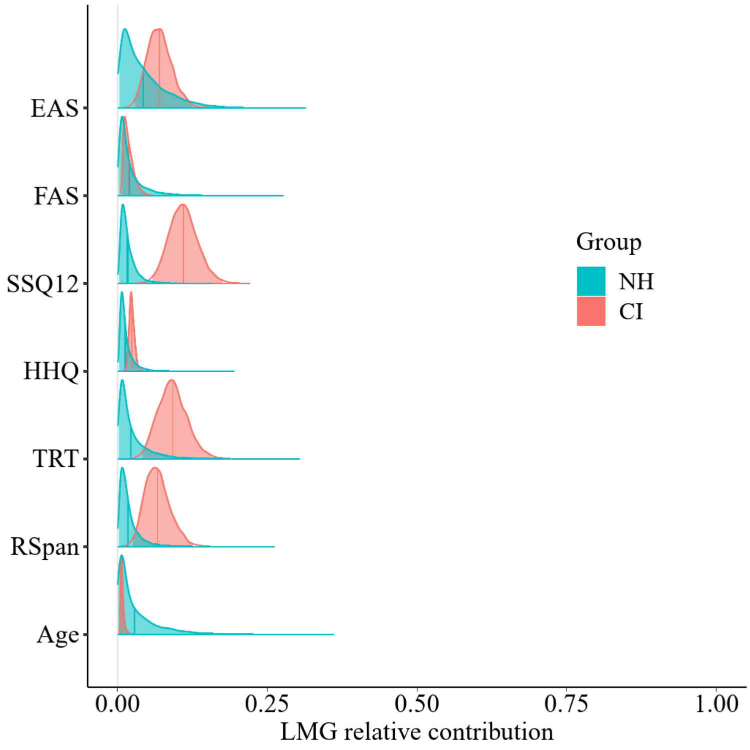


Figure 2.14. *Posterior relative contribution of predictor variables to individual-sentence differential drift rate by group. Predictor variables are age, hearing questionnaires (EAS, FAS, SSQ12 and HHQ), and cognitive tests (TRT and RSpan) scores.*

As can be seen in Figure 2.14, none of the variables were able to explain the variance in the NH group, whereas some seemed to explain some of the variability of CI users' listening efficiency. In particular, the SSQ12, EAS, and TRT scores accounted for 0.11, 0.07, and 0.09 of the variance, respectively. To explore any potential correlations with those variables, both plausible correlations and plausible population correlations (plausible p) for each of them were calculated following Ly and colleagues' analytic approach (Ly et al. 2017).

Moderate-to-weak correlations were found between CI users' scores in the SSQ12 (r=0.4), EAS (r=-0.3), and TRT (r=-0.4) and their performance on the behavioural task (listening efficiency). These, although not being strong relationships, suggest that the greater the hearing abilities of CI users (SSQ12 scores without reversing), the better their listening efficiency was in the behavioural task. Conversely, the more effort they reported in daily life (greater EAS scores) and the worse linguistic closure abilities they exhibited (greater TRT scores), the inferior their performance was in the behavioural task (left column plots in Figure 2.15). Although the 95% CrIs of the plausible population correlations (plausible p) just encompassed zero (SSQ12 [-.07, .64], EAS [-.61, 0.12], and TRT [-.65, .07]), there was reasonably strong evidence in favour of a correlation within the CI group. Specifically, there was 95.4% (p>0), 91.7% (p<0) and 95.1% (p<0) certainty that participants' listening efficiency correlated with SSQ12, EAS, and TRT, respectively (right column plots in Figure 2.15).

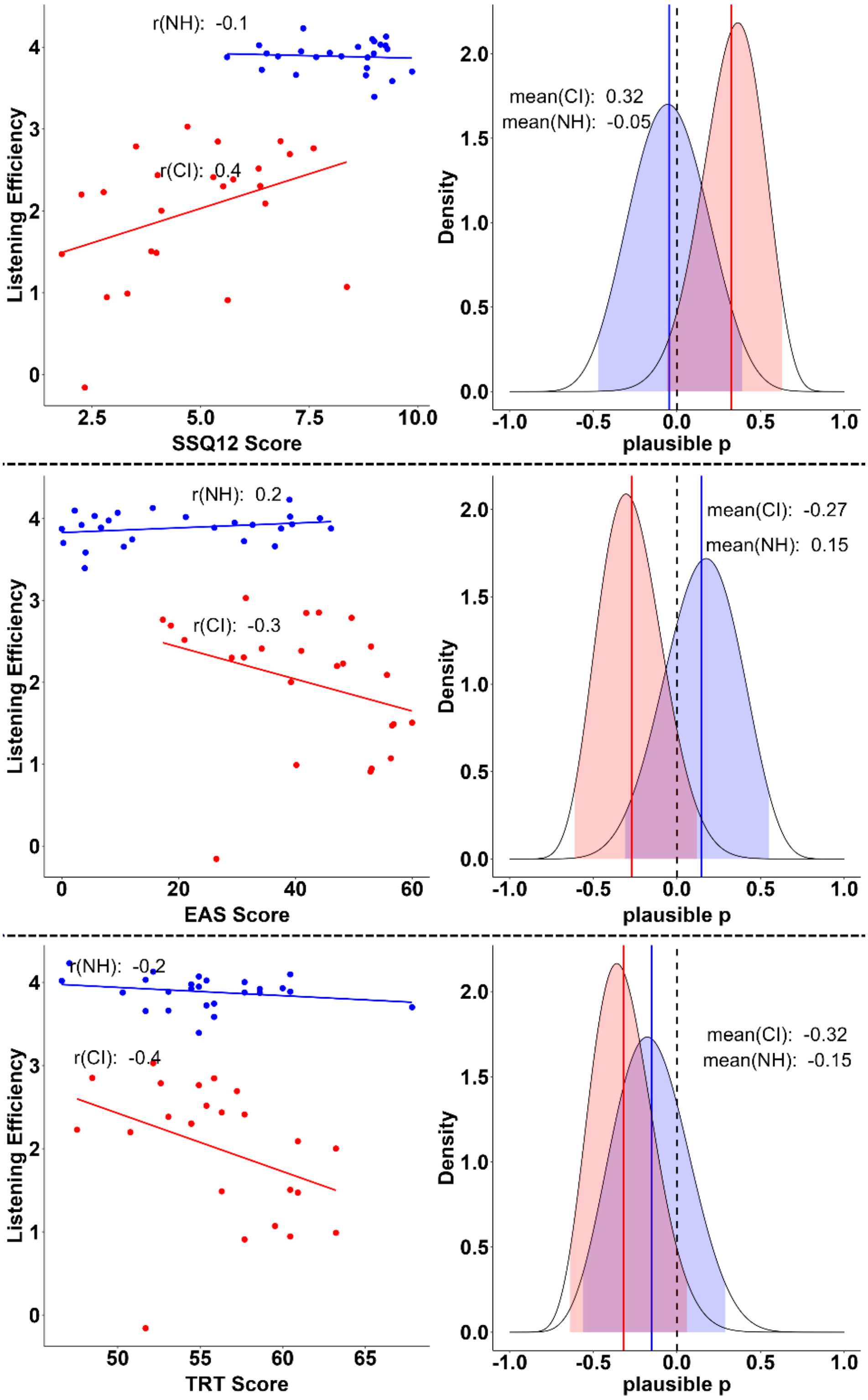


Figure 2.15. *Relationship between listening efficiency and SSQ12 (top panel), EAS (middle panel) and TRT (bottom panel) scores by group. Groups are plotted in red (CI) and blue (NH) colours. Plots on the left column display the posterior mean estimates of listening efficiency for each participant as a function of SSQ12, EAS, and TRT scores, respectively. Plots on the right column show the posterior distribution of the plausible population correlation (ρ) per each variable. In the later, coloured solid lines and shaded areas, represent the mean, and 95% credible intervals, respectively.*

The relationship between participants' subjective measures and their listening efficiency in the behavioural task were also explored. Figure 2.16 shows the partial contribution of these subjective measures (averaged across conditions) to the overall variability in listening efficiency.

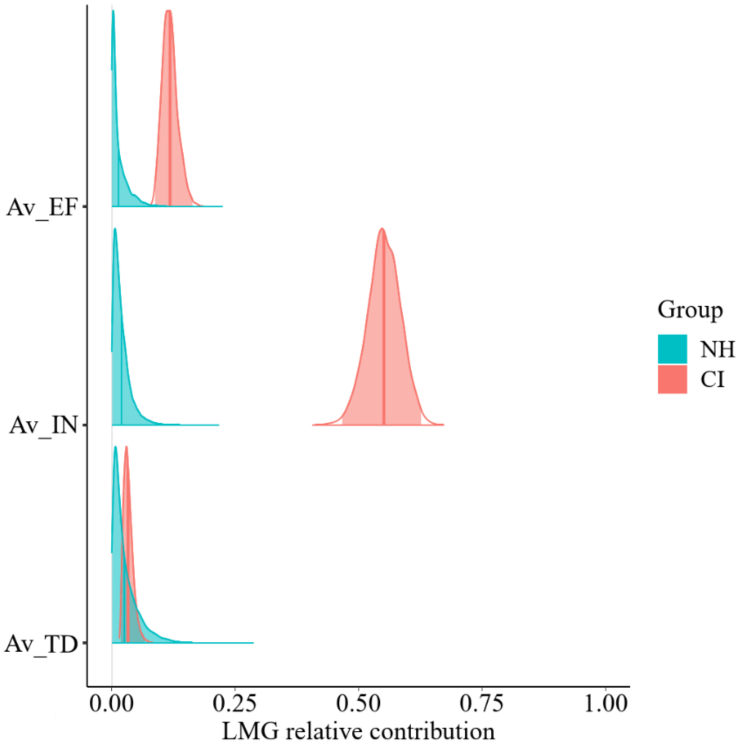


Figure 2.16. *Posterior relative contribution of task subjective measures (averaged across conditions) to individual-level listening efficiency by group. Predictor variables are participants' perceived listening effort (AV_EF), intelligibility (AV_IN), and task disengagement (AV_TD).*

Just like before, some task subjective measures seemed to act as individual predictors of listening efficiency only in the CI group. Both perceived intelligibility and effort contributed largely to explain CI users´ performance variability. Indeed, moderate and strong correlations were found between CI users' listening efficiency and their subjective scores of effort (r=-0.5) and intelligibility (r=0.9), respectively. Evidence of such correlations was also found at the population level (Figure 2.17) as revealed by 95% CrIs of the plausible population correlations that do not contain zero (AV_EF [-.72, -.06], AV_IN [.56,.91]).

These results suggest that CI users´ perceived effort and intelligibility reflected accurately their performance during the task.

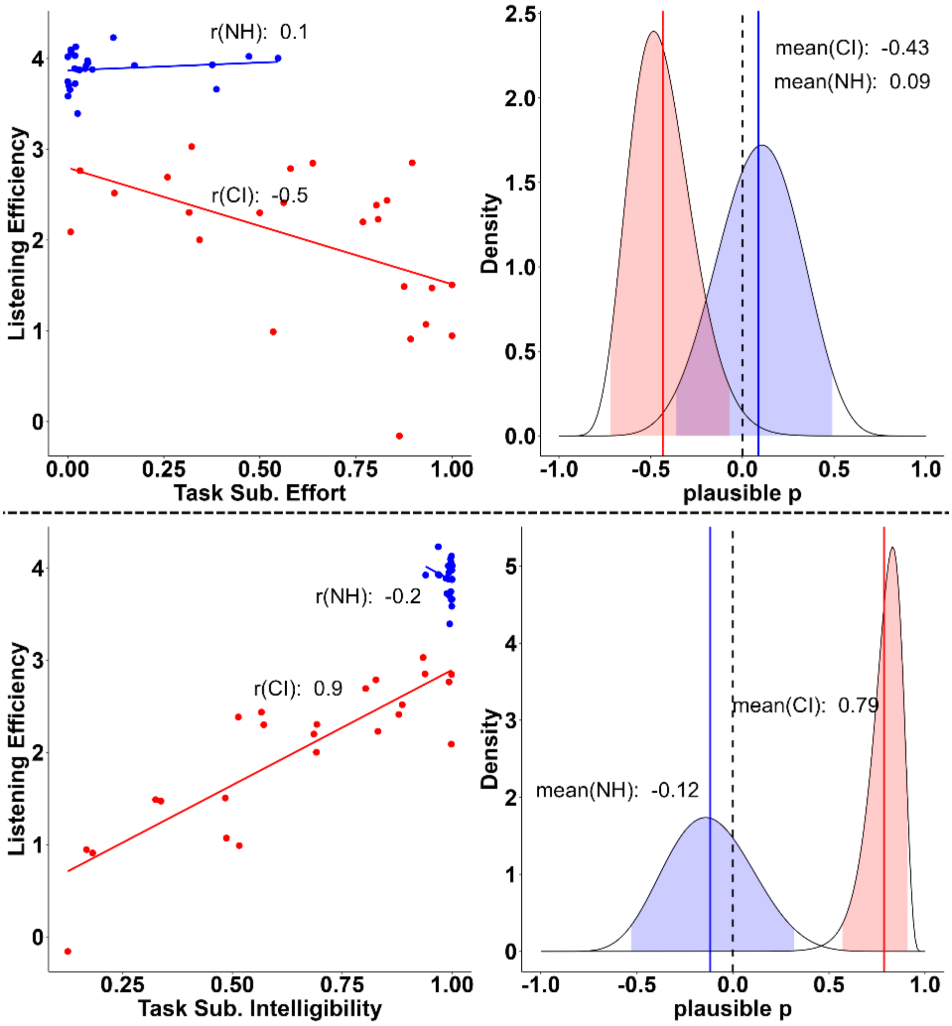


Figure 2.17. *Relationship between listening efficiency and both task subjective effort (top panel), and task subjective intelligibility (bottom panel) scores by group. Groups are plotted in red (CI) and blue (NH) colours. Plots on the left column display the posterior mean estimates of listening efficiency for each participant as a function of perceived effort and intelligibility, respectively. Plots on the right column show the posterior distribution of the plausible population correlation (ρ) per each variable. In the later, coloured solid lines and shaded areas, represent the mean and 95% credible intervals, respectively.*

2.6 DISCUSSION

In this chapter, a LBA model was used to perform a joint analysis of behavioural measures acquired in a laboratory experiment that aimed to assess the cognitive effort and listening performance of a group of CI users and a group of age-matched NH controls. The drift rate parameter was proposed as a putative metric of participants’ listening efficiency and its correlations with other LE measures (self-reported, task subjective, and cognitive) were examined following the plausible values approach.

CI users were disproportionately affected by moderate ecologically relevant levels of background noise

The between-group comparison revealed significant differences between CI users and NH controls as assessed by different measures of LE. Regarding self-reported daily life measures, EAS scores clearly showed that CI wearers reported considerably greater levels of listening effort in daily life than controls. This was consistent with their own perception of hearing abilities (SSQ12) and hearing handicap (HHQ), which again showed a clear disadvantage for CI users, compared with NH participants.

Participants were also consciously aware of their effort while performing the main task. CI users reported at least 50% more LE than their NH peers did in all experimental conditions. Such remarkable difference was apparent even in the easy condition (SNR: 20dB), despite task accuracy and, self-reported intelligibility, being near ceiling in both groups (approximately 90% and 100% in CI and NH groups, respectively). These results confirm the observation already made by many researchers that LE can be present even at ceiling levels of speech understanding performance (Pals et al. 2020; Pals, Sarampalis, and Başkent 2013; Winn et al. 2015; Winn and Teece 2021).

Such differences between groups were also confirmed by the behavioural results. The LBA´s drift rate parameter in sentence trials interpreted as a measure of listening efficiency showed a strong effect of group in all listening conditions.

Overall, CI users exhibited inferior listening efficiency (slower accumulation of evidence towards correct answers) during sentence trials compared with NH controls in all experimental conditions. That no effect of group was found in null trials suggests that the reduction in listening efficiency observed in the CI group during sentence trials must be associated with difficulties encountered during the listening process rather than the execution of the task itself, which was the same in both trial types. Such assumption is also supported by the non-decision time (t0) parameter of the LBA model and the cognitive tests' results. Both indicating that participants had similar cognitive abilities (working memory and linguistic closure) and spent the same amount of time to perceive/encode the stimulus and execute the task (button pressing). Therefore, the increased cognitive load perceived by CI users during the task (as expressed by their subjective scores) was consistent with and mirrored by their behavioural performance (listening efficiency).

Although listening was clearly more effortful for CI users than for NH participants, no differences in fatigue were found between groups. Similar scores were reported by participants in the FAS questionnaire and the MFQ, suggesting that participants overall experienced low levels of fatigue both in daily life and during task performance. These results support the fact that the experience of LE does not necessarily imply the presence of listening-related fatigue. Although it is reasonable to think that there must be a connection between both concepts (Hornsby et al. 2016), there is very little empirical support for a cause-and-effect relationship (McGarrigle et al. 2014b; Pichora-Fuller et al. 2016). Indeed, Alhanbali's (2017) study found low correlations between FAS and EAS scores and concluded that fatigue cannot be reliably predicted from self-reported effort.

Another common assumption is that fatigue could lead to task disengagement (Boksem and Tops 2008; Hockey 2013). In the present study our subjects reported low levels of fatigue, and thus, it is not surprising that the levels of disengagement were equally low in both groups. Indeed, previous studies have shown that once CI users engaged in communication they tend to persevere despite experiencing effortful listening (Eckert et al. 2016; Herrmann and Johnsrude 2020; Perea Pérez et al. 2022).

Drift rates from LBA models: a new metric of listening efficiency.

The speed of evidence accumulation (drift rates) towards correct answers (incorrect responses subtracted) was proposed as a measure of participants' listening efficiency. It was hypothesised that this metric would be able to capture differences in listening performance between groups and conditions. The results of the study confirmed this. As mentioned above, the listening efficiency metric was able to show a significant effect of group in all experimental conditions in sentence trials. Moreover, this measure was sensitive to changes in task demands, showing a declining trend as the task difficulty increased. This tendency was evident (as revealed by negative slopes across conditions) not only in the CI group but also in the control group. This reduction in listening efficiency among NH participants may not be too surprising considering their own subjective ratings during the task— they did perceive an increase in LE as the SNR worsened. This may prove the sensitivity of the drift rate parameter in capturing the SATO even when none of the behavioural measures alone could reflect such tendency (i.e. as revealed by similar accuracy and RT distributions of NH participants across conditions in Figure 2.11). In addition, the lack of effects in null trials supports the validity of the listening efficiency metric at reflecting changes in participants' performance and cognitive processing load during active listening.

Notably, listening efficiency was correlated with subjective measures of LE only in the CI group. The potential of an objective measure of listening performance that is consistent with self-perceived ratings of effort is very promising. Listening efficiency is not subject to individual bias and yet is able to reflect to some extend participants´perception of LE. This is an important quality, given that self-reported measures have been considered more sensitive than other methods to evaluate LE (particularly due to changes in task demand), and thus more relevant to audiological contexts (Francis and Love 2019; Johnson et al. 2015; Visentin et al. 2022). Nonetheless, further investigation is needed to examine the sensitivity of the listening efficiency metric and its relationship with self-reported measures of LE.

Although the speech recognition task used in the study was relatively simple for compatibility with the physiological measures simultaneously recorded, such simplicity is not required for the application of the proposed LBA analysis. Listening efficiency can be evaluated using any type of speech recognition task, as long as RT and accuracy are recorded. Therefore, the assessment of listening efficiency may also be relevant to clinical applications. This metric could provide a better evaluation of patients' performance, taking into account both the speech understanding ability and the cognitive load exerted. This approach could be applied to the tests batteries currently used in clinics. Indeed, speech in noise (SIN) tests, such as the Quick Speech in Noise Test (QuickSIN), Hearing in Noise Test (HINT), Bamford-Kowal-Bench SIN Test (BKB-SIN) and the City University of New York Sentences (CUNY) are commonly used to assess patients' intelligibility pre- and after implantation (BSA 2019). By additionally recording the response time, it would be possible to evaluate CI users' listening efficiency with the same test batteries already used by audiologists. Nonetheless, we should be mindful of the main limitation of this analysis– LBA models could be analytically costly. Certainly, RT decision models require a considerable post-processing analysis that, although common in research, may be less suitable in clinical environments due to time-constraints. Custom software should be developed to perform such analysis in "real time" and to provide an interpretation of the patients' performance with respect to the wider CI population. Although this poses a great challenge, the development of such software is feasible, just as other clinical solutions were provided in the past to evaluate otoacoustic emissions or auditory evoked potentials.

Individual predictors of listening efficiency

The plausible correlation analysis yielded significant associations between CI users' listening efficiency and their subjective effort and cognitive ratings, respectively. These findings suggest that better cognition and more positive self-reported listening experiences may be predictors of higher listening efficiency in CI users.

Indeed, although cognitive tests did not show any significant difference between groups, at the individual level, CI users with worse linguistic closure abilities (as measured by TRT test) exhibited poorer listening efficiency. Similar associations were already found by previous studies that concluded that the long-term memory process and lexical access abilities tapped by this test were correlated with increased speech in noise perception (Besser et al. 2013; Haumann et al. 2012; Kramer, Zekveld, and Houtgast 2009; Strand et al. 2018). Sabinne Haumann's study even considered the test to be a predictor of better postsurgical speech recognition performance in CI recipients. She suggested that the TRT test should be included in the CI candidacy criteria. The association between the TRT scores and listening efficiency, although moderate (plausible p = -0.3), is likely to be present in the CI population with 95% certainty, as revealed by the plausible correlation analysis.

Similarly, the scores of hearing questionnaires describing participants' hearing abilities (SSQ12) and perceived effort (EAS) in daily life were associated with their performance during the task. Higher listening efficiency was positively correlated with better hearing abilities and less perceived effort in everyday life. In the same way, CI users' self-reported measures of momentary effort and intelligibility during the task seemed to act as predictors of their listening efficiency. Participants were consciously aware of the level of speech understanding achieved, and thus, their subjective ratings accurately reflected their listening efficiency performance (plausible p= 0.8). Conversely, listening efficiency was inversely correlated with task perceived effort (plausible p= -0.4).

There is conflicting evidence concerning associations between behavioural and subjective measures of LE. While some studies, like the present one, have found correlations between participants' performance and their effort ratings (Koelewijn, de Kluiver, Barbara G Shinn-Cunningham, et al. 2015; Stenbäck et al. 2023; White and Langdon 2021), most of the research literature failed to find these associations (Alhanbali et al. 2019; Anderson Gosselin and Gagné 2011; Francis and Love 2019; Hornsby 2013; McGarrigle, Rakusen, and Mattys 2021; Shields et al. 2023; Strand et al. 2018). Although differences in the experimental design (different behavioural task and effort questionnaires) may explain in part the disparity in results, the main difference between previous studies and the present one is how correlations were calculated. Most studies only used one behavioural measure (either accuracy or response time) in their correlation analysis. However, the use of integrated behavioural measures is usually preferred when measuring cognitive or executive functions since they capture the SATO that usually gets lost using response time or accuracy separately (Bakun Emesh et al. 2021; Stafford et al. 2020). The listening efficiency metric used here, being an integrated measure that reflects intelligibility and effort, could have tapped into different domains of the listening effort construct, reflecting perhaps a combination of exerted and (self-) assessed effort(Francis and Love 2019).

Moreover, the Bayesian nature of the plausible correlation analysis could be more appropriate to explore associations between measures since it overcomes the limitations of multiple testing usually associated with p-values in the traditional frequentist approach. Although these could explain the associations found, at this point we can only speculate since there is not enough evidence to prove that this integrated measure provides a more comprehensive assessment of listening performance and thus favours correlations with other measures. More research is surely needed to explore the sensitivity of the listening efficiency metric and its relationship with other measures of LE.

Finally, the experimental design could have also contributed to improve the coherence and consistency across measures. For instance, the simultaneous acquisition of behavioural and task subjective measures is generally preferred to reduce within-subject variability across measures. Likewise, the ecologically relevant stimuli used in the study could also have favoured correlations with subjective measures, given that realistic stimuli are likely to evoke similar perceptions and reactions to those experienced in daily life (and reflected by self-reported questionnaires).

2.6.1 Limitations

It is known that effort measured in the laboratory is likely to differ from the effort experienced in the real world, particularly due to task differences (FUEL framework). Although the main behavioural task was designed to achieve some degree of ecological validity (by using meaningful sentences masked by realistic background noise at most frequent SNR levels), we could not replicate other aspects that play an important role when listening under naturalistic conditions. For instance, the presentation of the sound stimuli did not consider the spatial distribution of sound sources. Both the target speech and the background noise were played using a loudspeaker located in front of the participants. The use of different loudspeakers to present the target and the background noise, as well as a more appropriate distribution of them around participants could have produced a better immersive sound experience of a cafeteria environment. Moreover, no visual cues were available during the listening task. Many CI users rely on visual cues such as facial expressions and lip reading to enhance their speech understanding performance. Thus, the lack of visual cues could have affected the levels of LE compared to real life. Even so, this difference may have not been that substantial considering the correlations found between participants' self-reported measures in daily life and their listening efficiency scores during the task.

The SNRs at which participants performed the main listening task, although representative of everyday life sound scenarios for people with HL (Smeds et al. 2015; Wu et al. 2018), are not likely to pose any listening challenge for NH participants. This low level of task difficulty (positive SNR for all experimental conditions) could have contributed partially to the high listening efficiency exhibited by the NH group, and therefore the great disparity in results with respect to the patient group. The same reasoning could be used to explain partly the lack of correlations found between measures of LE within the NH group. One could assume that people with NH may expose themselves to more challenging listening environments in everyday life than the ones reproduced in the behavioural task.

2.7 CONCLUSIONS

A LBA model was used to perform a joint analysis of behavioural measures and assess listening efficiency in a group of CI and NH participants. The listening efficiency metric proposed here holds potential as a new outcome measure able to characterise the speed-accuracy trade-off (SATO) of participants' performance under challenging listening conditions. This metric was sensitive to changes in task demands and able to determine significant differences between groups. Under moderate ecologically relevant levels of background noise, CI users exhibited significantly inferior listening efficiency than their NH peers did in all experimental conditions. Only in the patient group, listening efficiency showed moderate-to-strong correlations with cognitive and self-reported measures of listening effort. These findings provide evidence of associations between objective and subjective measures of LE. This metric should warrant further consideration given its ability to assess both intelligibility and effort. Nonetheless, more research is needed to explore its sensitivity to changes in task demands and its consistency with other measures of LE under different experimental paradigms.

CHAPTER 3. PHYSIOLOGICAL MEASURES OF LISTENING EFFORT IN CI USERS AND THEIR CORRELATES WITH LISTENING EFFICIENCY

3.1 CHAPTER OVERVIEW

Physiological measures of LE are often chosen for their ability to provide objective evidence of instantaneous changes in central (brain activity) and/or autonomic nervous system activity due to cognitive processing load. Previous research using neuroimaging techniques has investigated the brain activity patterns of CI listeners during speech recognition. However, very little is known about the neural mechanisms that support speech comprehension under more complex naturalistic sound scenarios. In this chapter, we present the results of two physiological measures, fNIRS brain imaging and pupillometry, which were simultaneously recorded in a laboratory experiment to investigate the neural correlates of LE in the implanted population (compared with NH controls). Pupil results indicated that the exposure to moderate levels of background noise was highly taxing for CI listeners (but not for NH participants), which left them with less cognitive resources available to perform the listening task. Such lessening of cognitive resources in the CI group was reflected in task-evoked physiological measures that were reduced as the level of difficulty increased. As a result, task-evoked physiological measures were unable to detect differences in LE between the groups of participants. The combination of fNIRS and pupillometry was proved a feasible and complementary approach to investigate LE in CI users. Metrics from both physiological measures were associated with the listening efficiency scores of the CI group, suggesting that they may act as predictors of CI users’ listening performance. Overall, the study findings provided objective evidence that, regardless of good intelligibility, CI users may experience high levels of LE in real-world sound scenarios.

3.2 INTRODUCTION

Excessive LE is a fundamental challenge faced by many CI users (Alhanbali et al. 2017). The consequences may be severe: stress and fatigue associated with this sustained mental exertion may lead to social and professional withdrawal (Kramer et al. 2006; Nachtegaal et al. 2009), and could accelerate cognitive decline in older adults (Dryden et al. 2017; Lin et al. 2013). Despite substantial research over recent years, there is still much to be learnt about LE and its assessment, particularly in the implanted population.

Among all proposed ways of quantifying LE (section 1.5), physiological measures have the advantage of being objective and providing a "real time" evaluation of task performance. They can record changes in central and/or autonomic nervous system activity during effortful listening. Particularly, neuroimaging techniques have successfully been used to provide insights into the brain processes (central nervous system) involved during speech recognition (Eckert et al. 2016). However, the incompatibility of CIs with established neuroimaging techniques like fMRI made this assessment less accessible and challenging for CI listeners. Fortunately, the advent of fNIRS, a non-invasive optical brain-imaging technique, has offered a powerful means of safely studying cortical brain activity in the implanted population (Basura et al. 2018; Saliba et al. 2016). fNIRS has many advantages that make its use particularly suitable to auditory research. It is acoustically quiet, fully compatible with CIs, portable and suitable for use in unconstrained testing environments, and relatively robust to motion artifacts. Indeed, a growing body of research has used fNIRS to investigate the neural correlates of speech perception and LE in NH individuals (Lawrence et al. 2018; Rowland, Hartley, and Wiggins 2018; Wijayasiri et al. 2017), HA users (Rovetti et al. 2019), and CI listeners (McKay et al. 2016; Olds et al. 2016; Pollonini et al. 2014; Sherafati et al. 2022; Zhou et al. 2018).

Those studies testing CI users or using vocoders to simulate CI speech in NH participants showed some brain activity patterns that were associated with speech recognition in CI recipients. Some of the fronto-temporal areas identified were: i) the Superior Temporal Gyrus (STG), including auditory cortices, since they play a crucial role in the processing of auditory and visual speech information that can potentially predict CI outcomes (Anderson et al. 2019; Lawrence et al. 2018; Strelnikov et al. 2015); and ii) prefrontal areas, including left inferior frontal gyrus (LIFG), and dorsolateral prefrontal cortex (DLPFC), whose involvement is associated with adaptive control mechanisms that optimise speech recognition performance under effortful listening conditions (Lawrence et al. 2018; Sherafati et al. 2022; Wijayasiri et al. 2017; Wild et al. 2012).

Indeed, the activation of these areas in relation to LE has also been confirmed by fMRI (Davis and Johnsrude 2003; Wild et al. 2012; Zekveld et al. 2006, 2014), PET (Coez et al. 2014; Petersen et al. 2013; Strelnikov et al. 2015), and EEG studies (Dimitrijevic et al. 2019; Jiwani et al. 2016). However, it remains unclear whether the same brain activity patterns will be present when CI users listen to speech under more complex naturalistic scenarios such as the presence of background noise, to which CI recipients are especially susceptible (Badajoz-Davila, Buchholz, and Van-Hoesel 2020; Fu and Nogaki 2005).

There is very little research investigating the brain activity elicited by speech perception in real-world scenarios (Rowland et al. 2018), and even less testing actual CI users (Xiu et al. 2022). Nonetheless, the limited evidence available revealed no association between activation in frontal areas (e.g., LIFG) and LE (as measured by behavioural or subjective scores) in NH or CI listeners under naturalistic sound scenarios. Rowland and colleagues (Rowland et al. 2018) speculated that perhaps real-world sound scenarios pose a listening challenge that differs from other effortful situations (e.g., degraded or fast speech), and thus trigger different compensatory mechanism to support speech comprehension. This interpretation is also shared by other researchers (Peelle and Wingfield 2016) and deserves further investigation. They also suggested that combining fNIRS imaging with the simultaneous acquisition of other physiological markers such as pupil dilation would facilitate the understanding of the neural mechanisms involved in speech comprehension under naturalistic conditions.

Indeed, pupillometry is a physiological measure of the autonomic nervous system considered a putative marker of LE (McGarrigle et al. 2014b; Pichora-Fuller et al. 2016). Pupil dilation has been consistently used in hearing research to study changes in cognitive processing load during listening tasks (Naylor et al. 2018; Zekveld et al. 2018). Moreover, pupil responses follow the same activation pattern in relation to task performance as some aforementioned areas of the prefrontal cortex, such as the LIFG. Both show an inverted U-shape relationship with performance, so that greater responses occur to degraded-yet-intelligible speech (intermediate levels of difficulty) compared with both completely clear and unintelligible speech (Aston-Jones and Cohen 2005; Lawrence et al. 2018). This close correspondence of pupillometry and fNIRS imaging measures suggest that they may assess the same cognitive listening load (Zekveld et al. 2014). Indeed, the simultaneous acquisition of both physiological measures is likely to provide a more complete picture of the neural mechanisms involved in CI users' speech comprehension under challenging listening conditions. Such combination benefits from the high temporal resolution provided by pupillometry (60 to 1,000 Hz), which allows tracking rapid changes in cognitive load, while having the spatial resolution of fNIRS (approximately 2-3 cm), which offers insights about the brain areas that are additionally recruited in challenging listening conditions. Moreover, pupil responses can provide additional information about motivation and task engagement, which is crucial for a good interpretation of results.

Considering the advantages of combining both measures, we incorporated in our laboratory experiment the simultaneous acquisition of fNIRS-based brain imaging, targeting fronto-temporal cortical areas, and pupillometry to investigate the neural underpinnings of LE in the implanted population under naturalistic listening scenarios. For this purpose, a group of CI users and a group of age-matched NH controls completed a listening task while their brain activity and pupils were monitored. The task was designed to be representative of real-life sound environments. In particular, we used meaningful speech sentences masked by a continuous "cafeteria" background noise. The speech signals were treated to reflect the room reverberation conditions, and the levels of speech relative to background noise (SNR) that defined the three experimental conditions, were chosen to be representative of real-world listening environments (Smeds et al. 2015; Wu et al. 2018). The main hypotheses were that: i) CI users will show greater levels of LE compared to their NH peers in all experimental conditions, as revealed by increased activation in frontal brain regions and larger pupil dilations; and ii) this elevated effort in the patient (CI user) group will increase as a function of task demands, namely as SNRs worsen.

This chapter not only presents the analysis of pupillometry and brain imaging results, but also examines the physiological correlates of listening efficiency. Given that behavioural measures were also simultaneously recorded during the main task, we examined whether participants' physiological reactions were associated with their performance during the task. Therefore, correlations between listening efficiency scores (calculated in CHAPTER 2) and both physiological measures were explored.

3.3 METHODS

Study methods are described in CHAPTER 2. Nonetheless, this section provides information that is specifically relevant to pupillometry and brain imaging measures. See Figure 2.2 for a reminder of the experimental design.

3.3.1 Visual Stimuli

At the beginning of the experiment, a dark screen (approximately 0.2 cd/m2) was displayed, followed by a white bright screen (103 cd/m2), each exposure lasting for a period of 10 seconds.

During the main task, the screen displayed a grey colour background at a constant brightness level of approximately 30 cd/m2, which was intended to elicit intermediate levels of pupil size. Grey colour probe words were presented overlaid on top of this background at approximately 46 cd/m2. Once participants gave their answers, "yes" or "no", the selected option changed to turquoise colour of equal luminance. A fixation grey dot at 40 cd/m2 was also present throughout the experiment when no instructions/questions were displayed. It served as a reference for participants to look towards the centre of the screen.

3.3.2 Equipment

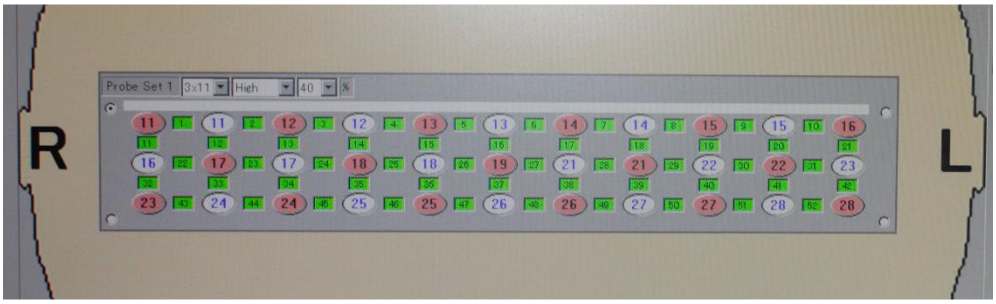
The main experimental task was conducted in a sound-attenuated room. Participants were seated at approximately 75 cm from a display screen with a loudspeaker (Model 8030A, Genelec, Iisalmi, Finland) mounted immediately above it. Participants entered their responses using a mouse and an “RTbox” button box.

A high lamp was placed at the back of the soundbooth to provide a minimum ambient lighting, so that participants could comfortably read the words displayed on the screen. This low luminance avoided interference with the fNIRS infrared light and was kept constant during the experiment to prevent luminance-related changes in participant’s pupil size. All other lights, including the main control room lighting were turned off to avoid distractions.

Brain activity was non-invasively measured using a Hitachi (Tokyo, Japan) ETG-4000 continuous-wave fNIRS system. The ETG-4000 measures simultaneously at wavelengths of 695 nm and 830 nm (sampling rate 10 Hz), and uses frequency modulation to minimize crosstalk between channels and wavelengths (Scholkmann et al. 2014). To obtain a good coverage of fronto-temporal brain regions a 3x11 optode array with a fixed source-detector spacing of 30 mm was used. This array is composed of 17 emitters and 16 detectors, which provides 52 measurement channels in total. To ensure a standardised array placement across participants, the international 10-20 system (Jasper 1958) was used to guide optode placement. The central optode in the bottom row of the array was placed on the forehead over position Fpz while the outermost optodes in the bottom row were aligned inferolaterally towards the preauricular points (in position T3/T4) (Figure 3.1). An optimal contact with the scalp was always sought, removing any hair interference when necessary. The final position of the optode array and the optode-scalp contact profile obtained were photographed as reference for each participant in both experimental runs. Figure 3.2 shows an example of the optode-scalp contact profile. The fNIRS equipment was placed behind participants, separated by a dense sound-absorbing screen to attenuate fan noise from the fNIRS system.

Pupil diameter was measured simultaneously during the experiment using a binocular Pupil Labs head-mounted eye-tracking system with a resolution of 400x400 pixels (per eye) at 60Hz sampling frequency. The pupillometry system had an additional “world camera” (1280x720 resolution at 60Hz sampling frequency) that was used for surface tracking to detect markers displayed on the display screen at the beginning and end of each experimental run. The eye-tracker cameras were adjusted to ensure an adequate field-of-view of each eye and of the display screen. Arm extenders were also mounted in the eye-tracker for a better adjustment of both eye cameras.

Figure 3.1. *Typical placement of the Pupil Labs eye-tracker and the fNIRS headset (optode array: 3x11) over participants’ head. Photograph of typical equipment setup on a volunteer’s head.*



Probe Set 1 | 3x11 | High | 40 %

11	1	11	2	12	3	12	4	13	5	13	6	14	7	14	8	15	9	15	10	16
11		12		13		14		15		16		17		18		19		20		21
16	22	17	23	17	24	18	25	18	26	19	27	21	28	21	29	22	30	22	31	23
32		33		34		35		36		37		38		39		40		41		42
23	43	24	44	24	45	25	46	25	47	26	48	26	49	27	50	27	51	28	52	28

Figure 3.2. *An example of the optode-scalp contact profile provided by the Auto Gain feature of the Hitachi ETG-4000 fNIRS system. The example shown displays an optimal optode-scalp contact as indicated by the green channels.*

3.3.3 Data acquisition

3.3.3.1 Brain regions of interest (ROI)

The ROIs defined in the study and the corresponding channels that cover these areas were:

- Right Dorsolateral Prefrontal Cortex (RDLPFC): channels 14, 24 and 25.
- Left Inferior Frontal Gyrus (LIFG): channels 30, 40 and 5.
- Auditory regions in the Superior Temporal Gyrus (STG): channels 42 and 52 in the left hemisphere and 32 and 43 in the right hemisphere.

The selection of channels covering each region of interest was based on the cortical sensitivity profile of the optode array, which has been calculated across the entire probe-set in previous studies conducted in our laboratory (Rowland et al. 2018; Wijayasiri et al. 2017).

3.3.3.2 Pupil data

Pupil recordings were continuously taken for the length of an entire experimental run (approximately 15 minutes). At the beginning of each run, the pupillary response range for each participant was measured following the method described by Piquado and colleagues (Piquado et al. 2010). Participants were presented with a completely dark screen followed by a bright screen, each lasting 10s. The difference in asymptotic pupil size elicited by these contrasting light intensities was taken as a measure of participants' pupil dynamic range and considered during the analysis. The surface tracker plugin was enabled and pupil markers were used to obtain the exact timestamp of every run's start and end time. Participants were overall instructed to look at a fixation dot located on the centre of the presentation screen, as long as they were not performing any other task that required their attention. Although de-blinking algorithms and other quality controls were performed during the pre-processing analysis, a real-time monitoring of participant's eyes were carried out by the experimenter to check the quality of the collected pupil data. After the experiment, these recordings (for each participant and run) were exported to and processed in the Pupil Player software, where an offline 2D model pupil detection algorithm was performed. Due to the large variability in participants' pupil size, a broad maximum and minimum range of 20 to 150 pixels were set in the detection algorithm to ensure that participants' pupil size was always detected.

3.4 ANALYSIS

3.4.1 Data pre-processing

Custom scripts were developed in Matlab (MATLAB R2018b, The MathWorks Inc., Natick, MA, USA) to carry out the pre-processing of brain imaging and pupillometry data.

3.4.1.1 Brain imaging

fNIRS data were analysed using HOMER2 package functions (Huppert et al. 2009) together with custom scripts developed in our laboratory for previous studies (Anderson et al. 2019; Lawrence et al. 2018; Mushtaq et al. 2019; Rowland et al. 2018; Wiggins et al. 2016).

Firstly, the raw fNIRS intensity signals were converted into changes in optical density, which then were filtered, using the *hmrMotionCorrectionWavelet* function, to correct for any motion artefacts. This wavelet filtering approach uses an exclusion threshold to eliminate outlying wavelet coefficients that are considered to be motion artefact. Coefficients laying more than 0.719 times the interquartile range below the first or above the third quartiles were removed. This criterion corresponds to the threshold typically used ($\alpha = 0.1$) in fNIRS motion artifact correction methods (Brigadoi et al. 2014; Cooper et al. 2012) when Gaussian distribution of wavelet coefficients is assumed.

Cardiac oscillations and low-frequency drift were also removed by bandpass filtering the optical density signals between 0.01 and 0.5 Hz. Then, a final conversion was applied to estimate relative changes in haemoglobin concentrations (HbO and HbR) using the modified Beer-Lambert Law (Huppert et al. 2009) with a differential path-length factor of six as the default value for both wavelengths.

To isolate the functional component of the haemodynamic signal from systemic physiological interference, the haemodynamic modality separation (HMS) algorithm (Yamada, Umeyama, and Matsuda 2012) was applied. It assumes that positive correlations between changes in HbO and HbR concentrations relate to changes elicited by systemic physiological oscillations and head movements whereas negative correlations are related to functional cerebral responses. As demonstrated in our previous work, the application of this algorithm has been shown to be beneficial, at a group level, to improve the reliability of fNIRS responses (Wiggins et al. 2016).

Finally, channels with poor optode-scalp contact were excluded using the scalp coupling index (SCI) method described by Pollonini et al. (Pollonini et al. 2014). In a compromise between the quality and quantity of the preserved data, a SCI threshold of ≥0.2 was applied leading to an exclusion of the worst 12.2% channels of the dataset and preservation of the remaining channels for statistical analyses.

3.4.1.2 Pupillometry

After identifying the time sync markers that inform about the experiment duration, the pupil data were split into two parts: the first 20 seconds that corresponds to pupil reactivity and the rest of the signal that constitutes the main trace data.

Firstly, a deblinking process was applied to the entire pupil dataset using a custom developed graphical user interface (GUI). This app allowed visualising pupil data for each participant’s eyes and run, and choosing the blink detection algorithm and settings (e.g., pre and post blink exclusion time) that best suits each particular case. It should be noted that participant’s pupil data were presented anonymously, in pseudorandom order to ensure a blind and objective deblinking process. Three deblinking algorithms were available: i) the ‘confidence’ method that uses the confidence parameter provided by Pupil Player to exclude data whose detection quality is less than a threshold set; ii) the “outlier” algorithm that classifies samples lying outside of a defined threshold range as blinks; and iii) the “gradient” method that establishes a time threshold above which rapid changes in pupil size are considered blinks.

Following this blinded manual tuning of the deblinking parameters to suit the characteristics of individual datasets, 79% of the data were filtered using the confidence algorithm, 16% using the outlier method and 5% were processed with the gradient algorithm. Three participants’ right-eye data were excluded due to poor quality (~3% of the data).

Once blinks were removed, the pupil reactivity and the main trace data were pre-processed separately. The pupil reactivity data (Figure 3.3) was used to identify, for each participants’ eyes, the maximum and minimum pupil diameter in pixels evoked by the initial dark and bright screens displayed during the first 20 s of each experimental run. This was done by defining a generic exponential function model to be fitted to the pupil reactivity data using the *nlilnfit* Matlab function. The model defined was a three degree freedom exponential function (*@(b,x) b(1) + b(2) * exp(-b(3)*(x-min(x)))*) with initial beta coefficients [100, -20, 0.5] and [80, 20, 0.5] for the dark and bright screen windows, respectively. Some sanity checks were performed immediately after the exponential fitting to confirm that the asymptotic values were correctly estimated within participants’ dilation range. Otherwise, whenever possible, missing values from one eye were calculated based on those from the other eye, as long as a positive linear correlation between both eyes´ pupil dilation ($r \geq 0.75$ & $p \leq 0.05$) existed across the full run. These maximum and minimum values of pupil reactivity were used to define participants’ pupil dynamic range. As expected, the dynamic range varied greatly across participants (minimum value range: 34-101, maximum value range: 55-165), but was consistent within-subjects across runs (SD: 3).

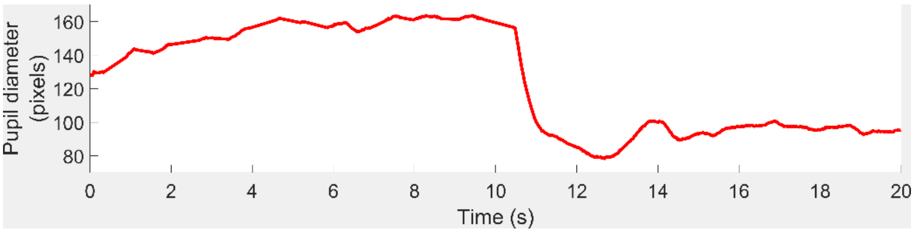


***Figure 3.3**. An example of pupil reactivity data (deblinked) recorded from participant NH_06. The first 10 seconds correspond to the pupil dilation evoked by the initial dark screen, whereas the last 10 seconds correspond to the pupil size measured during the bright screen exposure. The pupil dynamic range (minimum, maximum) estimated (by exponential fitting) for this participant was [94, 157] pixels.*

The main trace pupil data, after being smoothed with a low-pass filter (5 Hz cut-off frequency), were normalised with respect to each participants' pupil dynamic range. Such normalisation (Piquado et al. 2010) was applied to account for any age-dependent (Winn et al. 2018) and hearing-related (Koelewijn et al. 2017; Zekveld et al. 2018) changes in pupil responsivity, which allows the comparison of all participants' pupil data regardless of their age and hearing status. Such normalisation is especially important in the present study given that we are comparing individuals with very different hearing abilities (NH vs CI listeners). As a result of this normalisation, part of the data from nine participants were excluded due to missing dynamic range values. A total of 7.8% of the pupil data was dismissed. However, usable pupil data from all forty-nine participants were considered in the analysis.

3.4.2 Statistical Analysis

The analysis was carried out in Matlab to estimate single-subject responses to each physiological measure. Then, these metrics were exported into RStudio (Version 4.2.0; R Core Team, 2022) to perform statistical analysis.

3.4.2.1 Brain imaging response amplitude

To quantify the haemodynamic response amplitude on a channel-wise basis, general linear model (GLM) analysis was performed following the approach described in our laboratory's prior studies (Lawrence et al. 2018, 2021; Wijayasiri et al. 2017). A set of three regressors for each condition (corresponding to the canonical haemodynamic response plus its first two temporal derivatives) were considered in the design matrix. The inclusion of the temporal derivative terms provides more flexibility for the model to capture small differences in the delay and/or duration of the observed haemodynamic response with respect to the canonical response. Each individual trial was modelled as an epoch corresponding to the stimulation duration. For each condition, the canonical and temporal-derivative regressors were serially orthogonalised with respect to one another (Calhoun et al. 2004). Model estimation was performed using a two-stage ordinary least squares procedure (Plichta et al. 2007), including a correction that accounts for serial correlation (Cochrane and Orcutt 1949). The "derivative-boost" technique (Calhoun et al. 2004; Steffener et al. 2010) was used to estimate the response amplitude in a way that would be minimally affected by any differences in response latency or dispersion between conditions. Significant cortical activation (beta values for the canonical response function) was tested per each experimental condition (Easy, Medium, and Hard) on the "sentence vs null" contrast (trial type) at a group level. Channel-wise, two-tailed t-tests were used to assess significant activation on the "sentence vs null" contrast for each condition. Due to multiple channel comparisons, a false discovery rate (FDR) correction was applied using the original algorithm (Benjamini and Hochberg 1995) which assumes positive dependency among channels (Singh and Dan 2006). Single-subject average contrast values (sentence vs null trials) across the channels that constitute each ROI were extracted per each experimental condition and used as input in the Bayesian analysis.

3.4.2.2 Pupillometry metrics

The stimulus-evoked pupil data was analysed per each trial in a 6 seconds window relative to stimulus onset. Pupil baseline (BL) was estimated as the average dilation elicited during a second before the stimulus onset, when only the background noise was present. It was also used to perform baseline corrections to other pupillometry metrics. Baseline correction in this experimental design allowed the isolation of the stimulus-evoked pupil response from the dilation elicited by the background noise that was continuously present throughout the experiment. Once baseline corrections were applied, the following metrics were calculated: i) the mean pupil dilation (MPD) that was estimated per trial as the area under the curve divided by the duration of the analysis window; ii) the peak pupil dilation (PPD) calculated as the maximum dilation that occurred in each trial and expressed in percentage point of participants' dynamic range; and iii) the slope (SL) in which pupil data returned to baseline. The latter was fitted to the "above-baseline" pupil response component using the *polyfit* Matlab function in a time window that started 2 seconds after the trial onset (once the peak dilation occurred) and finished when the pupil dilation returned to baseline level. The time window was lengthened until the end of the trial (6 seconds) should participants' pupil dilation not reach the baseline. Given that the peak dilation was expected to vary among experimental conditions (easy, medium, and hard), a peak amplitude normalisation between zero and one was performed prior to the slope calculation. This amplitude normalisation allowed comparing slopes across conditions.

These metrics, except the slope, were calculated per each participant and condition at the trial level, and then averaged across eyes, blocks, and runs. The slope calculation, however, was done directly on the average pupil data to provide a more robust slope fitting. All the pupillometry metrics were then exported into RStudio for statistical analysis.

3.4.2.3 Bayesian inference

Differences in pupil dilation and brain activity responses between both groups of participants were examined using the R package *brms* (Bürkner 2017) that implements Bayesian multilevel models in Stan. These metrics were assessed on the interaction "group per condition" considering participants' random effects by group (e.g., *LIFG ~ Condition * Group + (1 | gr(Participant, by = Group))*). Student's t-distribution was set as the default prior family since it performs robust linear regression that is less influenced by outliers. Posterior distributions were estimated using MCMC (Van Ravenzwaaij et al. 2018) algorithms, over four separate chains (at convergence, Rhat=1), each with 2000 warmup iterations followed by another 2000 post-warmup iterations. Posterior predictive checks served as visual inspections of an adequate fitting between the model predictions and data.

The model conditional effects, predicted means, and 95% credible intervals were reported per group and condition. When relevant, population level effects were also calculated as the average and trend (slope) across conditions by group.

3.4.2.4 Plausible correlations

Plausible values correlation analysis (Ly et al. 2017) was performed in RStudio to examine relationships between participants' listening efficiency (integrated behavioural metric described in CHAPTER 2) and their pupillometry and brain activity responses during the main task. Firstly, the relative importance of pupil and fNIRS metrics as predictor variables of listening efficiency, was calculated in a multiple linear regression model using the Relaimpo R package (Grömping 2006). Then, the distribution of plausible population correlations (Ly et al. 2018) was calculated with respect to those physiological metrics that could best explain listening efficiency's variance. The plausible values approach allows generalising correlations from our sample of participants to the general population. Groups' mean correlation and 95% credible interval were reported. Reliable correlations were considered when the 95% credible interval did not contain zero or when high degree of certainty (>90%) suggested that the true population correlation (plausible p) was different from zero.

3.5 RESULTS

3.5.1 fNIRS results

As expected, the statistical analysis of all participants' brain activity in the contrast, sentence versus null trials, revealed significant activation in cortical areas typically recruited during speech in noise tasks: bilateral activation in the auditory cortices, activation in left inferior frontal region and right dorsolateral prefrontal cortex (Figure 3.4). Interestingly, significant deactivation during sentence trials (contrasted against null trials) was also observed in the frontopolar cortex (FP: channels 37, 47, and 48). This area was included as a secondary ROI targeted by channels 36,37,38,47 and 48 (Rowland et al. 2018).

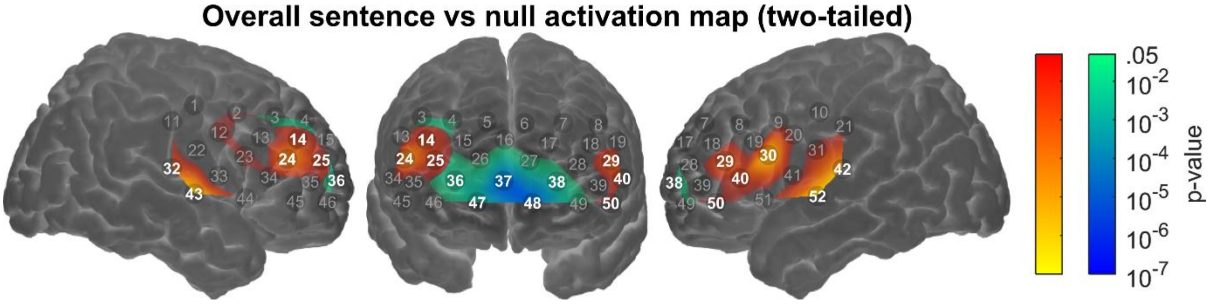


Figure 3.4. *All participants channel-wise statistical activation map in the contrast, sentence versus null trials. The channels that showed a significant result after FDR correction (q < .05) are highlighted. Red and blue areas indicate activation and deactivation, respectively, with gradient colour scale showing statistical significance (p-value).Note the maps are interpolated from single-channel results and the overlay on the cortical surface is for illustrative purposes only.*

3.5.1.1 Channel-wise group differences

The analysis per group showed that fNIRS amplitude responses presented greater variability among participants in the patient (CI) group compared to the NH group, which led to fewer channels reaching significant activation after correcting for multiple comparisons. As can be seen in Figure 3.5, only two channels overlying the left and right auditory cortices (channels 52 and 43) were significantly activated in the CI group. Nonetheless, CI users showed similar trends of activation and deactivation (although not statistically significant) to that exhibited by NH controls.

These activation patterns were present over the LIFG (channels 30 and 40) and RDLPFC (channel 24), as well as deactivation covering part of the FP cortex (channel 37).

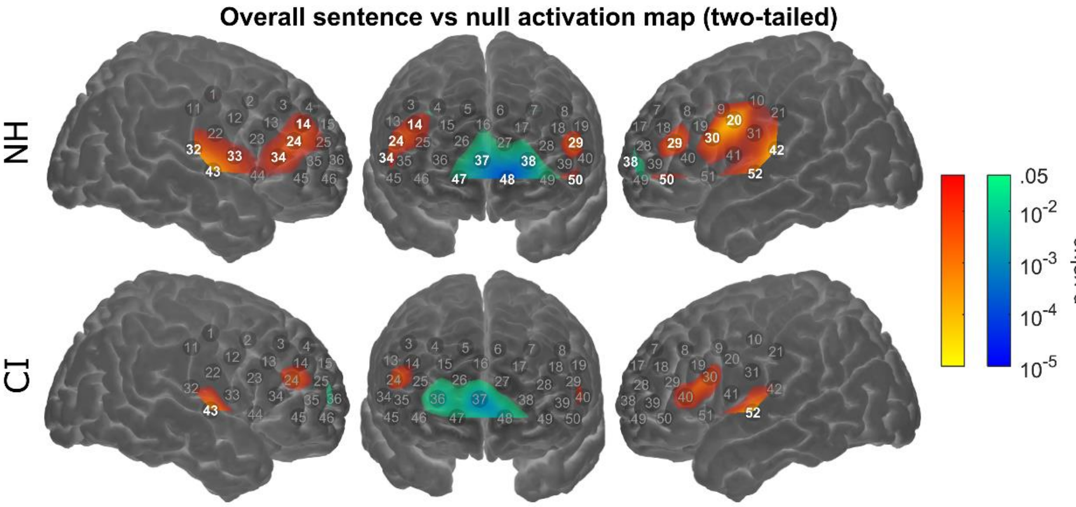


Figure 3.5. *Groups' statistical activation maps contrasted against null trials. NH and CI groups displayed at top and bottom panels, respectively. The channels that showed a significant result after FDR correction (q < .05) are highlighted. Red and blue areas indicate activation and deactivation, respectively, with gradient colour scale showing statistical significance (p-value). Note the maps are interpolated from single-channel results and the overlay on the cortical surface is for illustrative purposes only.*

3.5.1.2 Region of interest group differences

ROI statistical analyses were conducted to examine brain activity differences between groups across conditions. The model predicted means and 95% credible intervals of the a priori and secondary defined ROIs are shown in Figure 3.6. As can be seen, there was no main effect of group or condition in any ROIs as revealed by overlapping 95% credible intervals.

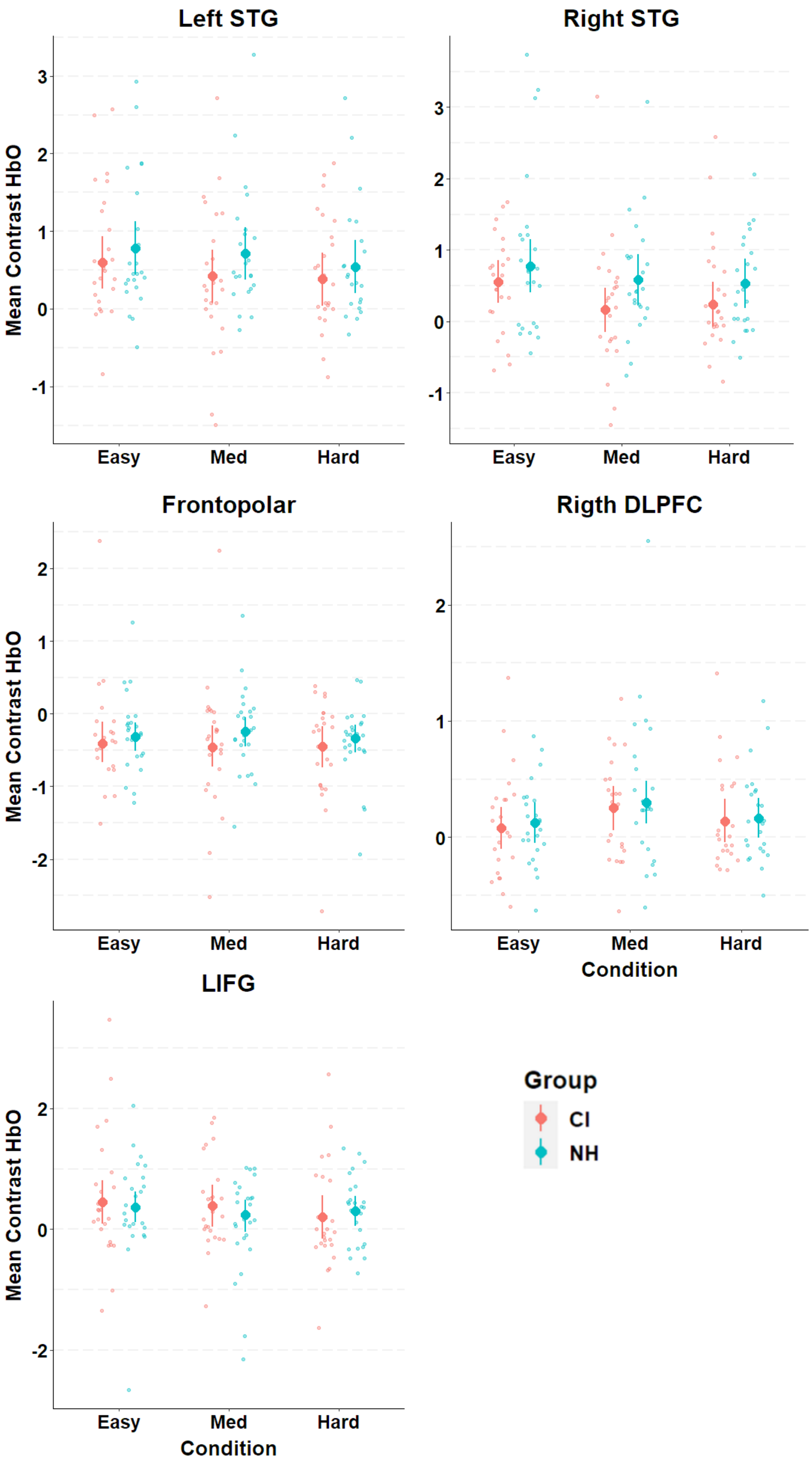


Figure 3.6. *fNIRS ROIs analysis. Model conditional effects over raw data for each ROI by group and condition (e.g., LIFG ~ Condition * Group + (1 | gr(Participant, by = Group))). The error bars display 95% credible intervals, the bold dots represent posterior means, and the small dots represent the raw data. ROIs are: left and right superior temporal gyrus (STG), right dorsolateral prefrontal cortex (DLPFC), left inferior frontal gyrus (LIFG), and frontopolar (FP). fNIRS response amplitude are shown as the mean contrast in oxyhaemoglobin (HbO).*

Overall, CI users showed lower mean fNIRS response amplitudes than their NH peers in all ROIs, with the exception of the LIFG where slightly greater amplitudes were shown in easy and medium experimental conditions. If we examine the LIFG haemodynamic time courses per condition, a decreasing trend of activation was observed in the CI group as the level of difficulty increased from easy to hard conditions (Figure 3.7). These results are contrary to our initial assumption in which greater activation was expected in the most demanding condition as an indication of higher cognitive load. Nonetheless, this tendency was not significant and thus no differences across conditions and/or groups were found.

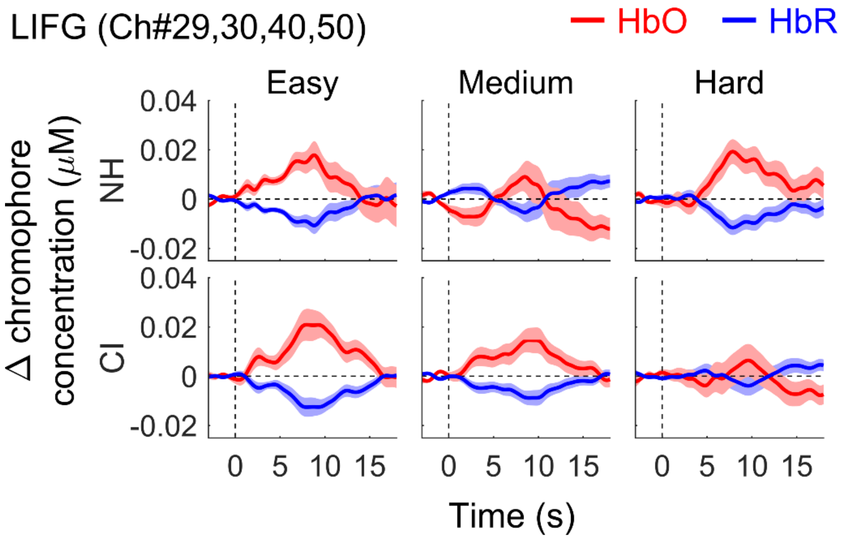


Figure 3.7. *Event-averaged haemodynamic time courses in the LIFG ROI per group. The red and blue traces show estimated changes in the concentration of oxyhaemoglobin (HbO) and deoxyhaemoglobin (HbR), respectively (average response with null trials subtracted out). The shaded area represents the 95% confidence interval around the group mean.*

The population level effects described in Table 3.1 also confirmed the lack of differences in the brain activity patterns exhibited by both groups. This can be seen by the posterior distribution of the between-group difference, whose 95% credible intervals contain zero in all ROIs and experimental conditions (the average and slope across conditions is also showed in Table 3.1).

Table 3.1. *Posterior group comparison in ROIs' fNIRs response amplitudes (population level effect). The mean and Highest Density Intervals (HDI) of the between-group difference are shown per each ROI. The HDI is interpreted as the 95% credible intervals (CrI) of posterior distributions.*

ROIs	**Between-Group Difference *(CI-NH)***		
	Mean	**95%CrI Lower**	**95%CrI Upper**
LSTG Cond. Easy	-0.19	-0.65	0.29
LSTG Cond. Med	-0.29	-0.79	0.17
LSTG Cond. Hard	-0.16	-0.65	0.32
LSTG Average Cond.	-0.21	-0.67	0.22
LSTG Slope Cond.	0.01	-0.16	0.18
RSTG Cond. Easy	-0.22	-0.71	0.27
RSTG Cond. Med	-0.42	-0.89	0.04
RSTG Cond. Hard	-0.29	-0.76	0.17
RSTG Average Cond.	-0.31	-0.64	0.03
RSTG Slope Cond.	-0.04	-0.32	0.27
LIFG Cond. Easy	0.08	-0.35	0.51
LIFG Cond. Med	0.16	-0.27	0.61
LIFG Cond. Hard	-0.11	-0.55	0.33
LIFG Average Cond.	0.04	-0.33	0.45
LIFG Slope Cond.	-0.09	-0.27	0.09
RDLPFC Cond. Easy	-0.04	-0.30	0.20
RDLPFC Cond. Med	-0.05	-0.31	0.22
RDLPFC Cond. Hard	-0.02	-0.27	0.23
RDLPFC Average Cond.	-0.04	-0.21	0.14
RDLPFC Slope Cond.	0.01	-0.15	0.17
FP Cond. Easy	-0.08	-0.43	0.24
FP Cond. Med	-0.21	-0.56	0.13
FP Cond. Hard	-0.12	-0.46	0.22
FP Average Cond.	-0.14	-0.42	0.13
FP Slope Cond.	-0.02	-0.19	0.15

Table rows represent each ROI per condition (Easy, Med, and Hard), the average across conditions, and tendency or slope across conditions. ROIs abbreviations are: left (LSTG) and right superior temporal gyrus (RSTG), right dorsolateral prefrontal cortex (RDLPFC), left inferior frontal gyrus (LIFG), and frontopolar (FP).

3.5.2 Pupillometry results

As expected, both groups of participants showed clear differences in their stimulus-evoked pupil dilations (baseline corrected) between sentence and control trials. Greater dilations occurred in sentence trials shortly after the sentence onset, reflecting participants' active listening and sentence perception (Figure 3.8). Conversely, as no speech sentence was played during null (control) trials, pupil dilation was triggered by the button-pressing task, which started approximately 2 seconds after the trial onset. Regarding group differences, it is noticeable that unlike NH participants, CI users' pupil dilations during null trials were, on average, inferior to baseline level. To examine participants' pupil responses during active listening only (excluding the task execution), all pupillometry metrics, except from pupil baseline that is independent of the trial type, were evaluated with control trials subtracted.

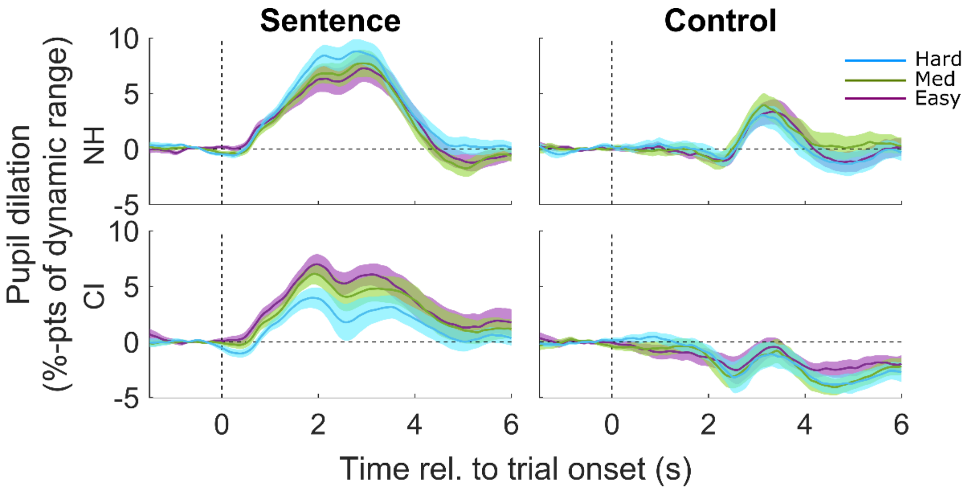


Figure 3.8. *Mean stimulus-evoked pupil dilations per conditions by trial types (Sentence, Control) and group (NH, CI). Pupillary response is plotted over the time course of 6 seconds relative to the stimulus onset. The three experimental conditions are represented by colour lines: purple (Easy), green (Med), and blue (Hard). Pupil dilation is baseline corrected and expressed as percent points of participants' pupil dynamic range.*

Figure 3.9 illustrates both groups' mean pupil response per condition, with null trials (control) subtracted. In general, NH participants showed a sharp inverted U-shaped curve, whereas CI listeners showed a more widespread inverted U-shaped curve response. As the level of difficulty increased across conditions, the steepness of the ‘peak’ response in the CI group became less pronounced, with flatter ‘tail ends’ returning to baseline. To examine these differences further, the following metrics (MPD, PPD, BL, SL) were statistically analysed to capture any significant differences between groups and conditions.

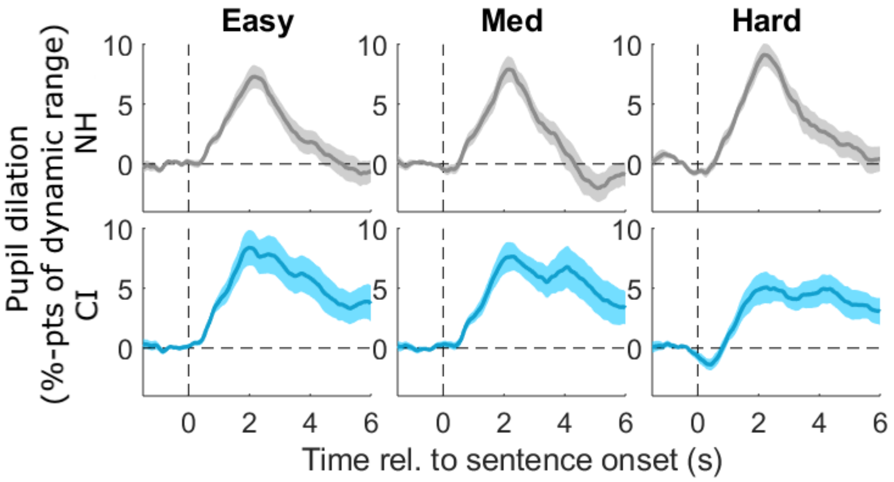


Figure 3.9. *Participants’ averaged stimulus-evoked pupil dilations (null trials subtracted) across conditions by group. Pupillary response is plotted over the time course of 6 seconds relative to the stimulus onset. The three experimental conditions (Easy, Med, and Hard) are displayed from left to right panels. Pupil dilation is baseline corrected and expressed as percent points (%-pts) of participants’ pupil dynamic range.*

3.5.2.1 Group differences in pupil metrics

The model predicted means and 95% credible intervals resulting from analysing the pupil metrics are shown in Figure 3.10. Both parameters, the MPD and the PPD, were able to reflect the same pupil reactivity pattern. CI users’ mean and peak pupil responses were greater than that showed by NH controls in Easy and Medium conditions, but not in the Hard one. Thus, CI users’ stimulus-evoked pupil dilation decreased as conditions became more difficult (MPD [Easy (M: 4.6), Medium (M: 4.4),Hard (M: 3.4)]; PPD [Easy (M: 4.8), Medium (M: 4.6),Hard (M: 3.3)]), as opposed to NH listeners who showed an increase in the hard condition (MPD [Easy (M: 2.6), Medium (M: 2.2),Hard (M: 3.4)]; PPD [Easy (M: 2.7), Medium (M: 2.1), Hard (M: 4.3)]). Nonetheless, these differences did not reach significance and therefore no effect of group and/or condition was observed in the MPD and PPD metrics.

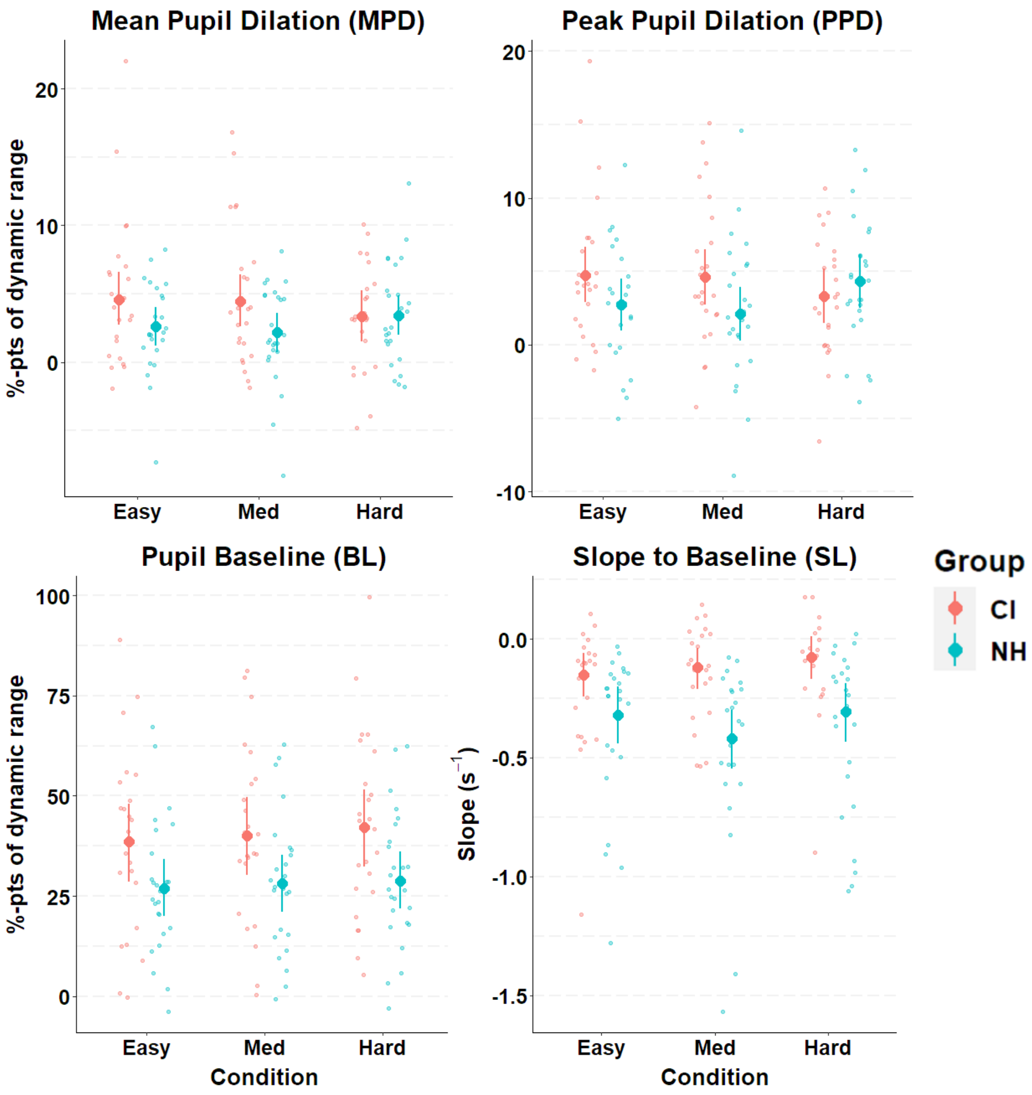


*__Figure 3.10.__ Model conditional effects over raw data for participants' pupil metrics by group (e.g., MPD ~ Condition * Group + (1 | gr(Participant, by = Group))). The error bars display 95% credible intervals, the bold dots represent posterior means, and the small dots represent the raw data. Pupil dilation is expressed as percent points of participants' pupil dynamic range.*

The analysis of the BL parameter however yielded significant group differences. This metric was analysed using the average across all trials since participants' pupil dilation at baseline was the same regardless of the trial type (less than 1% point of individual dynamic range difference between trial types). Overall, greater pupil dilations at baseline were shown by CI users compared with NH controls in all experimental conditions. These findings suggests that CI listeners already experienced a great deal of arousal due to the continuous background noise exposure. Moreover, such state of alertness exhibited by the patient group seemed to increase as the background noise level became louder. This upward trend of baseline pupil dilation across conditions was significant only in the CI group, as revealed by the positive mean of the slope across conditions and the 95% credible interval (Figure 3.11) that did not cross zero (M: 1.8, CrI:[0.3 3.2]).

Considering the population level effects, it is highly likely (97% probability) that this effect (CI users' BL > NH listeners' BL) would be observed in the wider population if tested under the same experimental conditions.

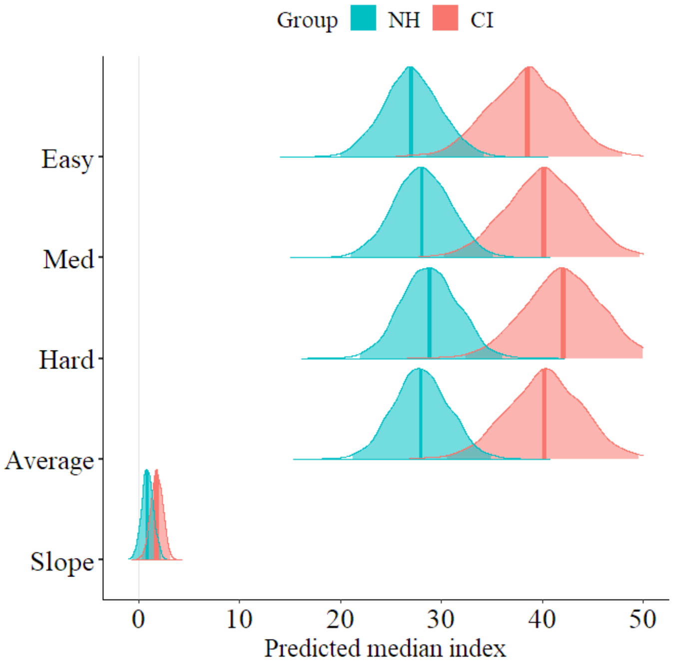


Figure 3.11. *BL Population effects per condition by group. Posterior group comparison of BL (averaged across trial types) per condition. From top to bottom, the first three rows represent the posterior distributions of BL in the three experimental conditions, whereas the last two rows display the average baseline dilation and slope (tendency) across conditions, respectively. Solid lines in posterior distributions represent the predicted median dilation expressed as percent points of participants' pupil dynamic range.*

Similarly, the speed at which participants' pupil size recovered to baseline at the end of the trials (SL) differed significantly between groups. Shallower slopes to baseline (i.e., slopes closer to zero) in the CI group suggests that CI listeners recovered from the arousal experience considerably slower than their NH peers during the trial. Moreover, the pupil recovery ramps in the CI group became flatter as the level of difficulty augmented across conditions. The effect of group is revealed by posterior distributions that hardly overlap between groups (Figure 3.12).

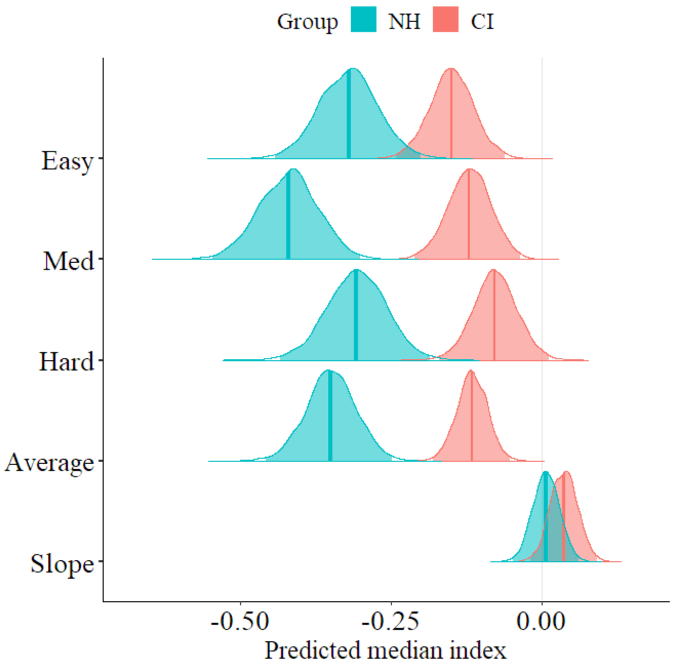


Figure 3.12. *SL Population effects per condition by group. Posterior group comparison of SL per condition. From top to bottom, the first three rows represent the posterior distributions of SL in the three experimental conditions, whereas the last two rows display the average and slope (tendency) across conditions, respectively. Solid lines in posterior distributions represent the predicted median slope (*s^{-1}*).*

3.5.3 Plausible correlations

Relationships between participants' listening efficiency scores (calculated in section 2.5.4) and their physiological reactions during the task, as recorded by pupil and brain imagining measures, were explored using the plausible values approach.

3.5.3.1 Correlations between listening efficiency and brain activity

A multiple linear regression model was used to examine whether fNIRs response amplitudes in any of the defined ROIs could act as predictors of participants' listening efficiency. The partial contribution of fNIRS metrics to the overall variability observed in participants' listening efficiency is shown in Figure 3.13.

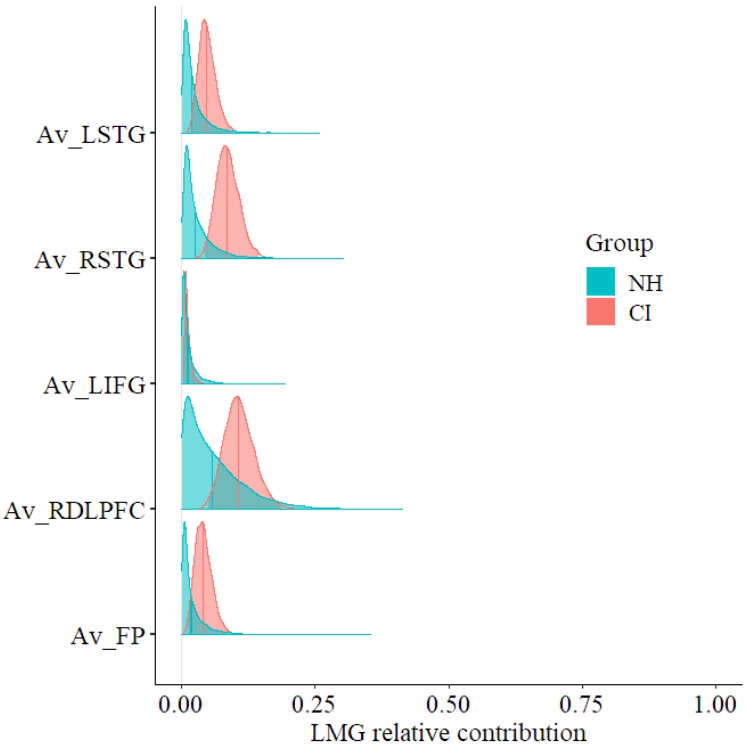


Figure 3.13. *Posterior relative contribution of fNIRS ROIs as predictor variables of listening efficiency by group. Predictor variables are the fNIRs response amplitudes averaged across conditions in all defined ROIs. ROIs are: left (LSTG) and right superior temporal gyrus (RSTG), right dorsolateral prefrontal cortex (RDLPFC), left inferior frontal gyrus (LIFG), and frontopolar (FP).*

Only in the CI group, the brain activity in the left and right auditory cortices (LSTG and RSTG), and the RDLPFC were able to account on average for 0.05, 0.09, and 0.1 of the variance, respectively. Indeed, greater listening efficiency was correlated with increased activation of the left (r=0.3) and right STG (r=0.4) in CI users. Although these were moderate associations, there is a high probability (92% and 97% for the left and right STG, respectively) of their prevalence in the population (plausible p>0) given the plausible population correlation's posterior distribution (LSTG plausible p: 0.3 [-0.1,0.6]; RSTG plausible p: 0.4 [-0.02,0.7]). Similarly, while no associations were found in the NH group, CI users seemed to exhibit a positive correlation between listening efficiency and their RDLPFC cortical activity (r=0.4). Although the 95% credible interval of the plausible population correlation just encompass zero (plausible p: 0.3, CrI: [-0.08, 0.6]), there is evidence (95% probability) that such association may exist (plausible p>0) within the population.

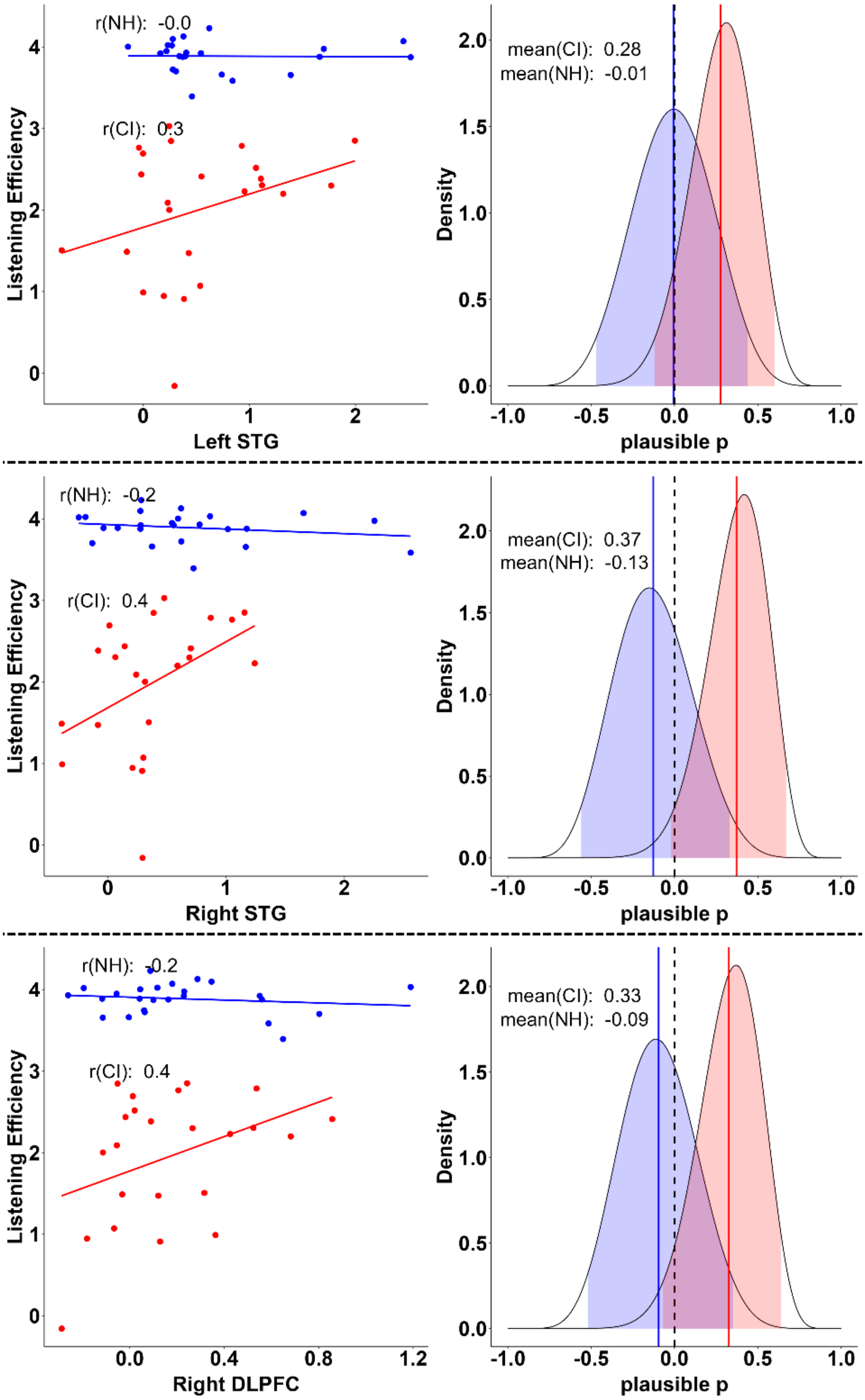


Figure 3.14. *Relationship between listening efficiency and the left (LSTG -top row), and right superior temporal gyrus (RSTG -middle row), and the right dorsolateral prefrontal cortex (RDLPFC- bottom row), by group. The CI and NH groups are displayed in red and blue colours, respectively. Plots on the left column show the posterior mean estimates of listening efficiency for each participant as a function of the LSTG, RSTG, and RDLPFC, respectively. Plots on the right column display the posterior distribution of the plausible population correlation (plausible ρ) per each variable. In the later, solid lines represent the mean, and shaded areas the 95% credible intervals.*

3.5.3.2 Correlations between listening efficiency and pupil metrics

Figure 3.15 shows the relative contribution of the pupil parameters considered in the analysis (averaged across conditions) as potential predictors of participants' listening efficiency. Only the BL and the SL to baseline metrics were able to explain some of the variability (on average 0.13 and 0.09 of the variance) present in the CI group's listening efficiency scores.

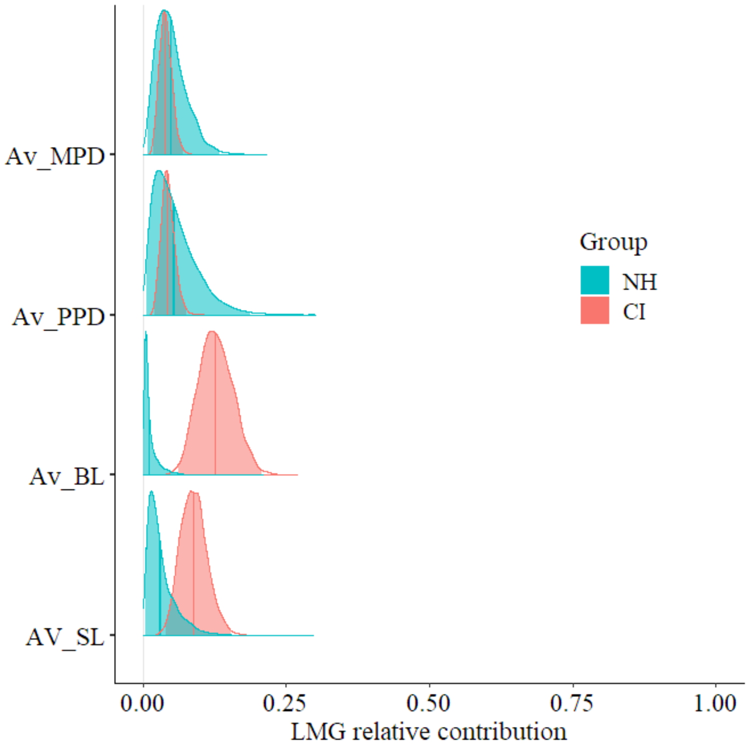


Figure 3.15. *Posterior relative contribution of pupil metrics as predictor variables of listening efficiency by group. All pupil metrics are averaged across conditions: mean pupil dilation (AV_MPD), peak pupil dilation (AV_PPD), baseline pupil dilation (AV_BL), and pupil slope to baseline (AV_SL).*

Hence, plausible population correlations (plausible p) were calculated between those metrics (BL and SL) and the listening efficiency parameter using Ly and colleagues' analytic approach. Analysis revealed moderate-to-weak correlations between CI users' pupil dilations and their listening efficiency scores while performing the main behavioural task (Figure 3.16). Particularly, greater pupil dilations at baseline were associated with better listening efficiency in the CI group only (r=0.4).

Such correlation is likely to be present in the population (plausible p>0) with a 96% probability as suggested by the plausible population correlation's posterior distribution (plausible p: 0.3, CrI: [-0.07, 0.6]). Likewise, flatter slopes to baseline were associated with worse listening performance in both groups. Although this correlation was relatively weak (r=-0.3), there was reasonably strong evidence in favour of its prevalence (plausible p<0) in the wider population, among both NH (84% probability) and CI listeners (91% probability).

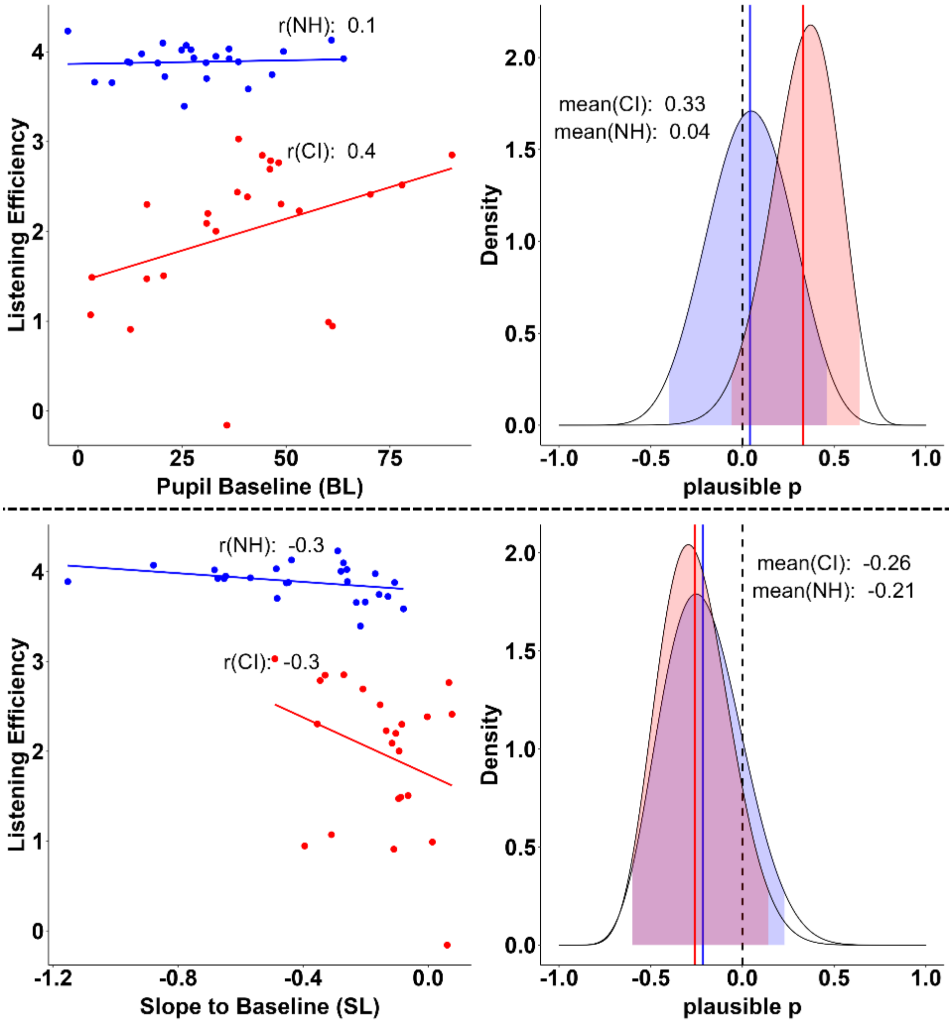


Figure 3.16. *Relationship between listening efficiency and both Pupil Baseline (top row), and Slope to Baseline (bottom row) by group. The CI and NH groups are displayed in red and blue colours, respectively. Plots on the left column show the posterior mean estimates of listening efficiency for each participant as a function of the BL and the SL, respectively. Plots on the right column display the posterior distribution of the plausible population correlation (plausible ρ) per each variable. In the later, solid lines represent the mean and shaded areas, the 95% credible intervals.*

3.6 DISCUSSION

This chapter presents the analysis of both physiological measures, pupillometry and brain imaging, acquired simultaneously in a laboratory experiment to examine the LE experienced by a group of CI users and a group of age-matched NH controls. Moreover, the study aimed to investigate the physiological correlates of listening efficiency, in other words, how physiological measures of LE may be associated to participants' listening efficiency.

Pupil dilations (BL and SL metrics) revealed increased LE in CI users compared with NH controls

Consistent with our first hypothesis, pupillometry was successful at detecting significant differences in the cognitive load experienced by the two groups of participants. However, unlike most studies that commonly use the task-evoked PPD to reflect changes in LE (Koelewijn et al. 2017; Naylor et al. 2018; Zekveld et al. 2014, 2018), in the current study the pupil BL and the SL to baseline were the only metrics able to confirm such effect of group. Indeed, the speed at which the stimulus-evoked pupil dilation returns to baseline (slope of recovery after peak) was significantly different between groups. While NH listeners showed a rapid and steeply return to baseline, CI users exhibited a more sustained pupil response pattern over time. These results indicate that CI users needed a prolonged time to recover from the arousal experienced during active listening. Moreover, such recovery period lengthened in time with increased task demands as revealed by flatter SL as conditions became more difficult (which confirmed hypothesis ii). These prolonged dilations to baseline have already been reported in previous studies and associated with aging (McGarrigle, Rakusen, et al. 2021), less efficient information processing and problem solving (Ahern and Beatty 1979; Bradshaw 1968), worse pitch discrimination(Bianchi et al. 2016), and worse resolution of speech understanding ambiguity (Winn and Teece 2020). Interestingly, we also found a correlation between participants' SL and their listening efficiency, so that more positive slopes (slower recovery to baseline) were associated with reduced performance (plausible p~ -0.3). This association was present not only in the CI group but also among NH participants.

Likewise, participant's pupil dilations at baseline differed considerably between the two groups of participants. Overall, CI users exhibited significantly greater pupil dilations than controls (in line with our first hypothesis) when they were listening to the background noise that was continuously present at baseline. The noise exposure itself turned out to be a taxing listening situation possibly due to the state of alertness required to identify any salient speech at the beginning of the trial, since participants did not know which trial type would come next. Previous studies has already shown that attention is key for CI listeners to process speech under challenging conditions (degraded speech simulated by vocoders)(Wijayasiri et al. 2017; Wild et al. 2012). Consistently with hypothesis ii), the allocation of cognitive resources needed for participants to maintain such focused attention increased as a function of the noise volume. The louder the background noise was, the larger their pupils became. This effect of condition was significant only among CI participants, who seemed to be very sensitive to the background noise. Although other factors (e.g., affective and emotional states, salience, and acoustic characteristics of the sound, etc.) could have contributed to the increase in CI users' pupil baseline, our interpretation of these high levels of arousal as reflecting cognitive processing load is consistent with participants' behavioural and subjective responses (simultaneously recorded) during the task. CI users showed less listening efficiency and increased perceived effort as the background noise became louder.

Reassuringly, this is not the only experiment in which this rise in pupil baseline has been observed in CI users. Russo et al.'s study (Russo et al. 2020) obtained similar pupillary responses from 10 CI patients while completing a speech audiometry in quiet and noise (+10dB SNR) conditions. The authors confirmed that the perception of background noise evoked an increase in CI users' pupil dilation prior to the task onset. Despite the paucity of studies investigating this phenomenon in actual CI listeners, these results seem to be consistent with previous research measuring pupillary responses in NH and HI participants (Aston-Jones and Cohen 2005; Koelewijn et al. 2017). They also found that the increase in task demands (e.g., tone discrimination, speech in noise task) induces the “anticipation” of cognitive resource allocation, which is reflected by larger pupil baseline.

Moreover, the LC tonic activity is believed to support performance, and thus, greater dilations at baseline (but below resource limits) have been associated with increased task accuracy (Alhanbali et al. 2020; Aston-Jones and Cohen 2005; Winn et al. 2018). Consistent with this hypothesis, we found that increase pupil dilations at baseline was correlated with greater listening efficiency (plausible p=0.3) only in the CI group. Thus, the analysis of BL dilations in realistic sound scenarios could be a predictor of listening efficiency in CI listeners.

Noise exposure lessened CI users’ cognitive capacity to perform the listening task

That noise exposure alone could be highly taxing for CI users has important implications. Given the assumption that individuals have limited cognitive capacity (Kahneman 1973; Pichora-Fuller et al. 2016); the more effort is exerted to cope with the increased state of alertness elicited by the background noise, the less cognitive resources would be left to be allocated to a listening task. Indeed, in Russo et al.’s study the task-evoked PPD (relative to baseline) was lower in the noise (5.7 % pupil change) compared to the quiet condition (6.5 %), possibly reflecting that in noisy environments CI users could not allocate to the task as many resources as in quiet. Our results seem to corroborate this observation too.

Although the analysis of the stimulus-evoked PPD and MPD (baseline corrected) did not show any effect of group, different trends of pupil responses across conditions were observed between the two groups of participants. While NH listeners showed the expected upward trend in PPD as the level of difficulty increased, CI users exhibited the opposite tendency. Their PPDs were slightly larger than that of their NH peers in easy and medium conditions, but smaller in the hard one. This declining trend in the PPDs showed by the CI group across conditions may indicate a lessening of attentional resources, given that high levels of arousal were already experienced at baseline. Similar results were reported by Barbara Ohlenforst and colleagues (Ohlenforst et al., 2017) who studied the performance and PPD of a group of NH and

HI individuals across a broad range of SNRs using two different masker types. They observed that when the listening condition became very difficult for HI individuals, their PPD that was generally greater than that showed by NH participants, dropped below that of the NH group. In Barbara Ohlenforst's study, this occurred at negative SNRs (approximately -4dB and -10dB for stationary and single talker maskers, respectively), when the HI group's intelligibility was very low (approximately 20% and 10% correct sentence recognition score for stationary and single talker maskers, respectively). In the present study the drop in CI users' PPDs happened at a positive SNR (+4dB) when their intelligibility remained relatively high (approximately 80% correct responses in sentence trials). The severity of participants' HL is likely to be the main cause (besides the type of background noise) of this shift towards more favourable SNRs in the threshold at which the decline in PPD happened. In fact, the HI group in Barbara Ohlenforst's study had mild-to-moderate HL, whereas the CI participants recruited in the present study had severe-to-profound HL, which inevitably reduces the range of SNRs at which they can listen to comfortably. Although the decline in PPD was interpreted by Ohlenforst and colleagues as a lack of participants' motivation and disengagement, here no evidence of disengagement was found. Indeed, high accuracy and longer reaction times observed in the behavioural task, as well as extremely low self-reported disengagement scores provided by the CI group (see section 2.5) suggest that participants kept on trying even during the hard condition. Although the direct comparison of both studies' results is not possible given the differences in the speech recognition tasks (button pressing versus sentence repetition), it is obvious that the decay in the PPD and MPD observed in the current study must be caused by reason(s) other than task difference. It possibly respond to participants reaching their maximum cognitive capacity, so that the remaining mental resources could have been insufficient to deal with the higher demands posed by the hard condition.

We are aware that this goes against the generally accepted idea that task-evoked changes in pupil size are independent from baseline dilation (Beatty 1982; McGarrigle, Rakusen, et al. 2021; Reilly et al. 2019). Although this may be the case under certain circumstances (e.g., when the pupil baseline is determined solely by the illuminance conditions), it might not be true for all experimental designs and subjects. Baseline dilation seems to be especially important in experiments whose condition of reference already involves a cognitively demanding task, like in the current experiment the presence of background noise. In these cases, the allocation of cognitive resources imposed by the task would depend on the current position at baseline on the effort curve (Duffy 1957; Yerkes and Dodson 1908). Such position depends on how much cognitive load is already experienced at baseline (which differs considerably between NH and CI listeners) and will determine the amount of remaining resources to be invested on the task until reaching resource limits (Granholm et al. 1996). Indeed, it is believed that when the baseline dilation reaches its maximum, it is suggestive of cognitive overload and task disengagement (Aston-Jones and Cohen 2005). The same reasoning could apply here, if all cognitive capacity is invested prior to the task (at baseline), the lack of remaining resources would result in task withdrawal. Considering this, it is recommended that baseline pupil dilations be examined to avoid bias when interpreting task-relevant pupil measures (Winn et al. 2018).

LIFG activity showed the same task-evoked cognitive load as pupil responses

Interestingly, similar results were found in the brain activity elicited in the LIFG. This cortical area, whose activation has been previously associated to effortful listening (Davis and Johnsrude 2003; Lawrence et al. 2018; Wild et al. 2012; Zekveld et al. 2006), showed a pattern of activation across conditions that mirrored the PPD and MPD results. A decreasing trend of activation across conditions was found in the CI group, with fNIRS response amplitudes that were on average inferior to that of the NH participants in the hard condition. Although these results were not statistically significant, the trends contradict our initial hypothesis that greater LIFG activation was expected in the CI compared to the NH group as a sign of higher cognitive load (especially in the most demanding condition). Given that functional brain imaging is evaluated in relative terms, in this case in the contrast sentence versus null trials, the fNIRS results presented here only reflect stimulus-evoked cortical responses. The activity elicited by the baseline noise exposure (which was common in both trial types) was not recorded by fNIRS measures. Therefore, the reasoning provided for the PPD and MPD pupil results could also be applicable here. If great LIFG activity already occurred during the background baseline, the lessening of cognitive resources available for the task would have been reflected by a reduced task-evoked LIFG activity in the hard condition. Although such conjecture is feasible and able to explain our results in coherence with the observed pupil reactions, the present study is unable to provide a baseline fNIRS measure to support it. Further investigation is needed to obtain evidence of any LIFG activity evoked by the noise exposure alone.

The simultaneous acquisition of fNIRS and pupillometry provide a more comprehensive assessment of LE

The combination of both techniques has been proved feasible to investigate LE in NH and CI listener under naturalistic sound scenarios. As expected, both physiological measures were able to reflect the same task-evoked cognitive load, as revealed by the same patterns of activation across conditions and groups, in both pupil metrics (PPD and MPD) and LIFG amplitudes. Moreover, they provided complementary information. The high levels of arousal experienced by CI listeners during the noise exposure that was not registered by fNIRS, was revealed by pupillometry (BL dilation). Therefore, the simultaneous acquisition of both measures provided a more comprehensive picture that facilitated the identification of the underlying cognitive load that seems to have compromised task resource allocation in the CI group.

CI users exhibited similar patterns of brain activity as NH controls

As expected based on the literature, the analysis of fNIRs responses showed significant activation in cortical areas typically recruited during speech in noise tasks (bilateral activation of the STG, LIFG, and RDLPFC). Regarding group differences, CI users showed on average lesser fNIRS response amplitudes in all ROIs (except for the LIFG as previously discussed) than NH controls. However, these differences did not reach statistical significance and therefore no effect of group was found in any ROIs. Although there is recent evidence of no significant differences in the auditory cortical activity between CI users and NH controls (Sherafati et al. 2022), the lack of differences in prefrontal areas is unexpected.

Indeed, the similarity of brain activity patterns between both groups of participants is surprising given that other measures of listening effort (subjective, behavioural, and pupil measures) did reveal a strong effect of group. There are several possible explanations for the observed fNIRS results. Perhaps the restoration of the hearing function following implantation made CI users to develop patterns of brain activity that increasingly resemble those of NH individuals (McKay et al. 2016). However, if it is so one would expect to find similar levels of intelligibility and/or LE in both groups, which was not the case. Another possible explanation is that the large variability found in listening efficiency scores among CI users could have also been reflected in a disparity of cortical activity (Olds et al. 2016). This would have prevented us from finding a common and representative pattern of CI users' brain activity.

The similarity of brain activity in both groups could also be attributed to the experimental approach used. The study design only allowed assessment of group differences in stimulus-evoked cortical activity, disregarding what was going on in participants’ brains during the background noise baseline. As previously discussed, if prefrontal activation responds to task demands in a similar manner to the LC activity measured by pupil responses (Zekveld et al. 2014), then the main difference in cognitive load between groups could have been evoked by the background noise. It seems feasible that we missed that effect of group given that previous studies have consistently found greater activity in the left prefrontal cortex (LIFG or Broca’s area) of experienced CI listeners (Petersen et al. 2013; Strelnikov et al. 2015) compared with NH controls (Jiwani et al. 2016; Sherafati et al. 2022; Zhou et al. 2018). However, all the aforereferenced studies involved tasks of speech recognition in quiet, which are considerably easier than listening in ecologically relevant levels of background noise. Therefore, it is not clear whether the same LIFG activity would occur under more challenging listening scenarios. In fact, Xiu et al.’s study (Xiu et al. 2022), could not find any change in relative alpha power at the LIFG as the background noise increased when CI users attended to natural speech from a television show. The authors suggested that factors such as the study design, and participants' mental state (motivation and fatigue) could act as confounders of CI users' neural activity. Moreover, listening to speech in real-world scenarios could recruit different brain areas as compensatory strategies to support speech comprehension (Adank 2012; Peelle and Wingfield 2016; Rowland et al. 2018; Wijayasiri et al. 2017). Further research is needed to elucidate the neural mechanisms that, under realistic sound scenarios, underlie effortful listening in CI users.

On the other hand, moderate relationships (plausible $p \approx 0.3$) were observed between CI users' fNIRS responses and their listening efficiency scores. We found that better listening efficiency was associated with increased bilateral activation in superior temporal areas. Although STG activity usually decreases during the post-implantation recovery period (as crossmodal plasticity is reversed), the speech understanding performance of experienced CI recipients seems to correlate with increased activity in auditory cortices. Indeed, previous research has shown the same positive relationship between activation in the auditory cortices and the speech recognition performance of NH (Binder et al. 2004; Lawrence et al. 2018, 2021; Mushtaq et al. 2021) and CI listeners (McKay et al. 2016; Olds et al. 2016). However, a recent fNIRS study found the exact opposite association, so that increase activation in the left STG (but not in the right) of CI users was negatively correlated with their speech test scores (Zhou et al. 2018). This inconsistency possibly relates to differences in CI users' speech understanding abilities. In fact, Cristen Olds and colleagues found that the positive correlation between the activity in the STG and speech perception was only present in CI listeners with good intelligibility while those with poor speech understanding performance exhibited large and indistinguishable cortical activations.

Likewise, only in the CI group, better listening efficiency was positively correlated with the strength of activation in the RDLPFC. Given that right frontal areas are generally involved in distractor suppression (Ptak 2012), this findings suggest that more efficient CI listeners implement greater inhibitory control to ignore irrelevant information (in this case the background noise) and direct their attention to task-relevant stimuli.

3.6.1 Limitations

The increase in pupil dilation at baseline exhibited by CI users as a function of background noise was interpreted here as reflecting anticipatory cognitive processing load, which is consistent with the results observed in the behavioural (listening efficiency) and subjective measures (perceived effort) simultaneously recorded during the experimental task (CHAPTER 2). However, it should be noted that we cannot discard the influence and potential contribution of other factors to CI participants' baseline pupil responses. In any case, the high levels of arousal experienced by CI users when listening to the background noise (compared to NH controls) were likely to affect their physiological responses during active listening (i.e. when speech sentences were presented on top of the background noise).

In this regard, the continuous presence of cafeteria background noise imposed an important limitation to the acquisition of physiological measures. Such experimental design seemed to elicit a sustained mental exertion in CI users (as revealed by high levels of arousal at baseline) that could not be evaluated by stimulus-evoked measures. Indeed, task–evoked physiological measures, being baseline corrected, were not able to show the entire picture of the cognitive load experienced by CI participants during the noise exposure. This mental effort possibly went unnoticed in the brain imaging results given that fNIRS measures (relative to null trials) only assessed participants' brain responses to speech sentences. Thanks to the simultaneous acquisition of pupil dilation as an additional physiological marker of LE, we could capture the pupil baseline reflecting the effect of the background noise. Modifying the experimental design to include a silent condition would facilitate the acquisition of fNIRS responses in the contrast silence versus noise. By doing so, it would be possible to confirm whether significant changes in cognitive load are evoked by the presence of background noise during active listening. Moreover, having a silent baseline would facilitate the direct comparison of pupil results with respect to other pupillometry studies whose baseline is usually measured in quiet.

Although 49 participants (25 NH and 24 CI users), is generally considered a good sample size to provide reliable group level fNIRS data (Wiggins et al. 2016), we are aware that the large variability found in speech recognition performance (as indicated by listening efficiency scores) may have increased variance in our CI group's brain-imaging data. Larger sample sizes grouped by performance level could overcome this limitation and provide evidence of any distinction in cortical activity among CI listeners as a function of intelligibility.

3.7 CONCLUSION

The study results suggest that CI users experienced higher cognitive load compared with NH controls, as revealed by larger pupil dilations when listening to speech under ecologically relevant levels of background noise ("cafeteria" background noise at +20, +10, and +4dB SNRs). CI users' LE increased as a function of task demands, namely as the background noise became louder. Indeed, the noise exposure alone may have required a state of alertness that could have depleted a great deal of CI users' cognitive capacity (as revealed by baseline dilation), leaving very little attentional resources to perform the listening task (as revealed by task-evoked PPD). Care should be taken when analysing task-evoked physiological measures in isolation as they may overlook any underlying cognitive processing (baseline) that can compromise task resource allocation. The combination of fNIRS brain imaging and pupillometry resulted in an appropriate approach to provide a more comprehensive picture of participants' cognitive listening load. Both physiological measures showed moderate associations with listening efficiency scores in the CI group, suggesting that those metrics may act as predictors of CI users' performance. These findings are particularly relevant to everyday life listening environments, and provide objective evidence of the experience of LE commonly reported by the CI community.

CHAPTER 4. PERCEIVED LISTENING DIFFICULTIES OF ADULT CI USERS UNDER MEASURES INTRODUCED TO COMBAT THE SPREAD OF COVID-19

Chapter adapted from: Perea Pérez F., et al. (2022). Perceived listening difficulties of adult cochlear-implant users under measures introduced to combat the spread of COVID-19. Trends in Hearing, 26, doi: 10.1177/23312165221087011

4.1 CHAPTER OVERVIEW

Following the outbreak of the COVID-19 pandemic, public-health measures introduced to stem the spread of the disease caused profound changes to patterns of daily-life communication. This chapter presents the results of an online survey conducted to document adult CI users' perceived listening difficulties under four communication scenarios commonly experienced during the pandemic, specifically when talking: with someone wearing a facemask, under social/physical distancing guidelines, via telephone, and via video call. Results from ninety-four respondents indicated that people considered their in-person listening experiences in some common everyday scenarios to have been significantly worsened by the introduction of mask-wearing and physical distancing. Participants reported experiencing an array of listening difficulties, including reduced speech intelligibility and increased listening effort, which resulted in many people actively avoiding certain communication scenarios at least some of the time. Participants also found listening effortful during remote communication, which became rapidly more prevalent following the outbreak of the pandemic. Potential solutions identified by participants to ease the burden of everyday listening with a CI may have applicability beyond the context of the COVID-19 pandemic. Specifically, the results emphasised the importance of visual cues, including lipreading and live speech-to-text transcriptions, to improve in-person and remote communication for people with a CI.

4.2 INTRODUCTION

The outbreak of the Coronavirus (COVID-19) pandemic in early 2020 profoundly changed patterns of daily life communication. The imposition of social-distancing measures, together with a rapid shift towards remote online communication methods, transformed social interactions with family and friends, access to essential services, and ways of working. Individuals with HL are thought to have been disproportionally affected by some of these developments (Chodosh, Weinstein, and Blustein 2020; Grote and Izagaren 2020; Ideas for Ears Ltd 2020; Maru et al. 2021; Naylor, Burke, and Holman 2020; Saunders, Jackson, and Visram 2020; Tavanai, Rouhbakhsh, and Roghani 2021; Ten Hulzen and Fabry 2020). Among people with HL, those who use a CI may have been particularly affected by the public-health measures introduced to combat the spread of COVID-19, because of the greater degree of HL (i.e., severe-to-profound HL) associated with this intervention. We sought to document, through an online survey, the perceived listening difficulties experienced by adult CI users during this unprecedented period, and to see whether transferable lessons could be learned to guide future research aimed at alleviating the challenges of listening with a CI.

We had a particular interest in probing participants' daily-life perceptions of LE during, compared with before, the COVID-19 pandemic, as well as possible sequelae of elevated perceived LE, such as listening-related fatigue and risk of social disengagement (McGarrigle et al. 2014a; Pichora-Fuller et al. 2016). Prior research has shown that listening to speech is more cognitively demanding for CI users than for people with NH (Perreau et al. 2017; Russo et al. 2020), even under favourable acoustical conditions (Pals et al. 2020; Winn et al. 2015). Indeed, pre-pandemic, CI listeners reported experiencing high levels of LE and listening-related fatigue in everyday life (Alhanbali et al. 2017; Hughes et al. 2018). This increased mental exertion may negatively affect people's ability to focus and sustain attention (Zekveld et al. 2011), as well as to retain important information in memory (McCoy et al. 2005; Tun, McCoy, and Wingfield 2009). Such difficulties may in turn impair communication success (Hetu et al. 1988; Wie et al. 2010), social participation (Barker, Leighton, and Ferguson 2017; Hughes et al. 2018; Kramer et al. 2006; Mick et al. 2014; Nachtegaal et al. 2009), long-term cognitive health (Lin et al. 2013; Pichora-Fuller et al. 2015) and overall quality of life (Carlsson et al. 2015; Chia et al. 2007; Dalton et al. 2003; McRackan, Hand, Cochlear Implant Quality of Life Development Consortium, et al. 2019).

Early reports, in the media and the scientific literature, suggested that the public-health measures introduced to combat the spread of SARS-CoV-2 (the virus responsible for the COVID-19 disease) had a disproportionate impact on people with HL (Chodosh et al., 2020; Grote & Izagaren, 2020; Ideas for Ears Ltd, 2020; Naylor et al., 2020; Saunders et al., 2020; Tagupa, 2020; see Tavanai et al., 2021, for a recent review). The use of facemasks, which became mandatory in many countries (on public transport, in healthcare settings, and in other public spaces), received particular attention. Studies showed that facemasks could hinder speech intelligibility because they muffle sounds and attenuate the voice (Goldin, Weinstein, and Shiman 2020; Magee et al. 2020; Ribeiro et al. 2020), and increase LE in the presence of background noise (Rahne et al. 2021). Moreover, masks create a visual barrier that obscures the speaker's mouth and lower part of the face. This was shown to further affect communication, especially among people with HL, who often rely on lipreading and facial cues to aid speech comprehension (Atcherson et al. 2017; Chodosh et al. 2020; Grote and Izagaren 2020; Naylor et al. 2020; Saunders et al. 2020; Ten Hulzen and Fabry 2020). Indeed, even experienced CI users with good overall proficiency in understanding speech still rely on visual cues to optimise communication performance in real-world listening situations (Moberly et al. 2020). The widespread use of facemasks was therefore expected to have a negative impact on communication for CI users especially.

Compounding the uptake of facemasks, social distancing rules also became established internationally after the COVID-19 outbreak, since increased physical distance between people was proven to reduce the risk of droplet transmission (Jones et al. 2020). However, the requirement to remain several metres apart (commonly two metres) could also have led to less favourable acoustical conditions, since, with a greater distance between conversational partners, the level of the target speech relative to background sound (the "target-to-background" or SNR) is reduced. Similarly, a greater distance leads to a reduction in the direct-to-reverberant ratio, which is an indicator of the level of the direct sound from talker to listener compared to the level of the reverberant sound that has reflected off a room's surfaces. Listening in noise is known to be more challenging for people with HL than for people with NH (Dimitrijevic et al. 2019; Koelewijn, de Kluiver, Barbara G. Shinn-Cunningham, et al. 2015; Needleman and Crandellt 1995; Pang et al. 2019; Shukla et al. 2020), and people with HL are especially sensitive to the deleterious effects of room reverberation on speech intelligibility (Badajoz-Davila et al. 2020; Eurich, Klenzner, and Oehler 2019; Hazrati and Loizou 2012; Kressner, Westermann, and Buchholz 2018). Physical distancing measures were therefore expected to have a further negative impact on communication for CI users.

At various stages throughout the COVID-19 pandemic, most people were obliged or advised to spend periods of time self-isolating in their home, whether to shield themselves from the virus or to reduce community transmission. Accordingly, the pandemic saw a rapid replacement of in-person interactions by remote communication. Healthcare services, for instance, in many cases underwent a rapid transition to telemedicine and virtual care during the pandemic (Bokolo 2020; Reay, Looi, and Keightley 2020; White et al. 2021). Many patients experienced a reduction of in-person visits to access primary care, mental-health counselling, and other health services, that increasingly switched to remote delivery. Working from home also became the "new normality" for many employees all over the world (Kniffin et al. 2021; Wang et al. 2021), with interactions with peers and colleagues relying almost exclusively on virtual online meetings. Even communication with family and friends took place predominantly online during the pandemic.

Virtual communication, especially online video calling, offers some advantages in terms of being able to control the acoustic environment during communication (e.g., adjustable volume, live captioning, visual indication of who is speaking), which could potentially benefit people with HL. A recent survey of 120 audiologists in the UK showed positive experiences of teleaudiology (Saunders and Roughley 2021) during the pandemic, nonetheless some concerns about poor internet connection and patients' technology familiarity were highlighted. Indeed, despite the advantages of video calls, previous studies showed that the increased reliance on remote communication may impose an additional burden on people with HL (Ideas For Ears 2018; Ideas for Ears Ltd 2020; Naylor et al. 2020; Tavanai et al. 2021). Naylor et al.'s study found that people with greater HL reported inferior hearing performance during video calls compared to in-person communication. Likewise, video calls and telephone calls were considered an issue for communication during the pandemic as reported by a survey on 249 respondents with HL (Ideas for Ears Ltd 2020).

It must be noted that, for many, living through the pandemic had a variety of consequences for health and wellbeing outside of listening challenges. Aside from potential long-term health effects of the virus itself, the lockdown and quarantine measures around the world imposed a forced social isolation that is associated with negative psychological effects. Brooks et al. (2020) reviewed the psychological impact of quarantine based on 24 studies from multiple countries including the United States of America, Canada, Sweden, Australia, Taiwan, and China. They concluded that the psychological impact of quarantine is wide ranging, substantial, and potentially long-lasting. Some of the psychological effects that have been reported include moderate-to-severe stress, anxiety, loneliness and depression (Brooks et al. 2020; Hyland et al. 2020; Razai et al. 2020; Wang et al. 2020). Individuals with HL or other communication disabilities may have been at increased risk of experiencing these psychological effects during the pandemic (Razai et al., 2020). Indeed, Naylor et al. (2020) concluded that COVID-related restrictions may have created an additional emotional burden that is stronger among people with greater HL. Moreover, it is plausible that the risk of social isolation that is already attributed to HL (Chia et al. 2007; Mick et al. 2014; Pronk et al. 2011; Shukla et al. 2020) may have been worsened as a result of the COVID-19 restrictions (Tagupa 2020).

To the best of our knowledge, no research has so far investigated the potential impact of COVID-19 public-health measures on CI users' everyday listening experiences, covering both in-person and remote social interactions. Nor has much attention been given to the perceived LE (and potential sequelae) associated with communicating under these measures. Therefore, we designed an online survey to investigate perceived listening difficulties of adult CI users under four commonly occurring communication scenarios during the pandemic, specifically when communicating: with someone wearing a facemask, under social/physical distancing guidelines (~2 m), via telephone, and via video call. Participants' listening experiences were examined based on six communication items (intelligibility, LE, need of repetition, disengagement, anxiety/stress, and listening-related fatigue), designed to probe both acute listening challenges and medium-term consequences. Where relevant, we asked whether participants' listening experiences during the pandemic were better or worse than they had been beforehand. We planned to perform comparisons within both in-person communication scenarios (facemask vs social distancing) and remote communication scenarios (telephone vs video call) to examine the importance of visual cues under these two modes of everyday communication. Finally, the survey sought CI users' views about strategies and technological solutions that may help to improve communication in in-person and remote scenarios. Results of the study could inform interventions and provide reliable advice to help people with severe-to-profound HL to communicate during these challenging times. Such lessons could also be applicable in post-pandemic society where online communication, for instance, may remain prevalent.

4.3 METHODS

The study was approved by the North West - Greater Manchester Central Research Ethics Committee (REC reference: 20/NW/0141).

4.3.1 Participants

The survey was aimed at adults aged 18 or over, who had at least one CI, spoke fluent English, had capacity to give informed consent and had no known cognitive impairments. Participation was voluntary and no incentives were offered. To access the survey, participants had to read the participant information sheet, confirm that they met the inclusion criteria as defined above, and provide informed consent. The survey took approximately twenty minutes to complete. However, there were no time restrictions and thus respondents could take as much time as they needed to answer all questions. A “previous” button was included throughout the survey allowing participants to go back and modify their answers if needed.

4.3.2 Distribution

The online survey was open for recruitment from July to October 2020. A link to the survey was emailed to all members of the NIHR Nottingham BRC participant database who met the inclusion criteria. The questionnaire was further disseminated by national and regional hearing charities and organisations in the United Kingdom including the Royal National Institute for Deaf People (https://rnid.org.uk/), the National Cochlear Implant Users Association (https://www.nciua.org.uk/) and Ideas for Ears (https://www.ideasforears.org.uk/). The survey was also publicised on NIHR Nottingham BRC social media feeds. It is estimated that approximately 320 CI users from the UK were invited to participate in the study. Nonetheless, it is unknown how many CI users could have seen the study advertisement in the organisations’ webpages, which are internationally accessible.

4.3.3 Survey development

The survey was designed to explore adult CI users’ perceived listening difficulties during in-person and remote communication under the measures introduced to combat the spread of COVID-19. The survey design was informed by validated questionnaires, such as the Speech, Spatial and Qualities of Hearing Scale (SSQ)(Gatehouse and Noble 2004) and the Effort Assessment Scale (EAS) (Alhanbali et al. 2017), that retrospectively evaluate respondents’ real-world listening experiences. However, given the unique context and purpose of our survey within the COVID-19 pandemic, we did not use, nor intend to develop, a standardised questionnaire in the present study.

The survey was implemented using the Jisc online survey platform (https://www.onlinesurveys.ac.uk/) and comprised 37 items in total (see Appendix A:

SURVEY ITEMS for a full reproduction of the survey items). Following an adaptive questioning procedure, some items (conditional questions) were only displayed where relevant according to a participant's prior responses (e.g., only participants who use a contralateral HA were asked about the frequency of HA use). Participants were required to answer all questions, with the exception of conditional and open (free-text) questions. The survey items were grouped into four sections: 1) demographic (age, gender, education, employment, and country of residence) and hearing information (hearing-device usage and experience, onset of hearing loss, and ways of communication in daily life); 2) measures affecting in-person communication; 3) remote communication; and 4) potential solutions to minimise any impact.

In the in-person communication section, we asked separately about the impact of two public-health measures introduced to control the spread of COVID-19: the use of face masks and the imposition of social/physical distancing (based on the instruction in the United Kingdom to keep at least 2 metres away from others, a widely adopted rule at the time the survey was conducted). Please note that as the recommended distance changed over the course of the pandemic, participants were instructed to answer the questions considering their overall experiences of having to maintain a minimum distance from others. In the remote communication section, we asked about participants' experiences using two modes of remote communication: telephone and video calls.

Both sections 2 and 3 followed a similar structure. Firstly, participants were asked to evaluate their current listening experiences during the COVID-19 pandemic (e.g., Q10.1. “For each question below, please select the option that best reflects your experience in this or these situation(s)”). Secondly, participants were asked about how their listening experiences have changed since the introduction of COVID-19 related measures (e.g., Q10.2. “Considering your listening experiences before and after the COVID-19 outbreak, how much do you think your communications have changed due to the speaker wearing a face mask?”). Thirdly, they rated which specific issues were causing them difficulty in a certain communication scenario (e.g., Q11. “The following is a list of potential challenges associated with listening to someone who is wearing a facemask. Please rate how relevant they are according to your experience"). Finally, they reported the degree to which they were avoiding certain communication scenarios because of adverse listening experiences (e.g., Q12. "How often do you find yourself avoiding face-to-face communication because of difficulty hearing someone who is wearing a face mask or covering?").

In total, six communication items were used to assess participants' listening experiences: intelligibility ("how much of the person's speech are you able to understand?"); effort ("how much mental effort do you have to put in to achieve this level of understanding?"); need of repetition ("how often do you ask the speaker to repeat (part of) the message?"); disengagement ("how often do you give up trying to communicate because the effort required was too great?"); anxiety/stress ("did you experience any feelings of anxiety or stress as a result of difficulty communicating?"); and fatigue ("to what extent did the communication leave you feeling tired/fatigued?").

For most survey items, responses were given on a five-point scale with appropriate labels as anchors at the endpoints. For example, for the questions enquiring about listening effort, the endpoint anchors were "no effort" and "lots of effort". Survey items enquiring about frequency of occurrence used five-point scales with category labels "never", "rarely", "sometimes", "often", and "almost always". Similar five-point scales are used in validated questionnaires commonly employed in the literature to assess self-reported fatigue (Fatigue Assessment Scale) (Michielsen et al. 2004) and hearing handicap (Hearing Handicap Questionnaire) (Gatehouse and Noble 2004). Survey items enquiring about changes in perceived listening difficulties from before to during the pandemic used five-point Likert scales with labels "much less", "less", "no difference", "more", and "much more".

A small number of open questions (free-text answers) were also included to collect: i) additional details about participants' hearing devices; ii) participants' listening experiences; and iii) potential solutions to improve daily-life communication. In the final section, participants rated a list of potential solutions according to the extent that they felt they might benefit from each.

Seven members of the Patient and Public Involvement group of the NIHR Nottingham Biomedical Research Centre (BRC) reviewed and provided feedback on the content and technical functionality of the survey during development.

4.3.4 Analysis

Ninety-four responses in total were coded and exported from the Jisc online survey system into a Microsoft Excel spreadsheet. The response data were anonymised, and participants were identified by a unique code. Descriptive statistics and analyses were performed using R software (Version 3.6.2; R Core Team, 2019). Non-parametric paired comparisons (Wilcoxon signed-rank test with continuity correction) were performed to test for differences between measures in the same category (e.g., facemask use vs. social distancing), while single-sample Wilcoxon signed-rank tests (with continuity correction) were used to test for significant changes from before to during the COVID-19 pandemic. The Holm method (Holm 1979) was applied to account for multiple comparisons across the full set of tests performed in this study. All p-values reported in the text are the corrected values, meaning that they can be compared against a conventional $p < .05$ threshold for statistical significance.

Exploratory factor analysis (EFA) was conducted using the ‘psych’ package (Revelle 2020) to determine the number of underlying constructs assessed by the six communication items. The non-graphical Cattel’s scree test (Cattell 1966) from ‘nFactors’ package was used to determine the number of factors to retain. An ordinary least squares estimation procedure was used to find the minimum residual (minres) solution using the ‘fa’ function. Only items with factor loadings of at least 0.5 were considered good indicators for the underlying factor. Bootstrapped confidence intervals for factor loadings were calculated with one thousand iterations. Scores on the “intelligibility” item were reversed prior to factor analysis, so that greater scores would in all cases reflect a worse listening experience.

Participants' (optional) responses to the three open (free-text) questions were analysed using a simple descriptive approach, with themes and categories selected based on Elo and Kyngäs's guidelines for inductive content analysis (Elo and Kyngäs 2008). Text responses were coded based on keywords (e.g. facemask, social distancing) to form the main themes that were then grouped into categories (e.g. difficult to lip-read). The number of mentions and example statements for each identified theme and category were also provided. Please note that participants' text responses were analysed for informational purposes and that the study conclusions were based solely on quantitative results.

4.4 RESULTS

4.4.1 Participant demographics and hearing profile

Ninety-four participants completed the survey, 73 women, and 21 men. Most participants were older adults, with the modal age category being 70-79 years old (Figure 4.1). Participants spanned age categories from 30-39 to 80+. Most participants were UK residents (92%), currently retired (56%), and with higher or postgraduate level of education (57%). Only seven international participants accessed the survey from the United States, Australia, Ireland, and Portugal. On average, the onset of HL in the implanted ear occurred at age 30-40, although 16% of participants had been deaf since birth or within the first year of life, and 8% lost their hearing later in life (over 60 years old). Most participants had more than 10 years' experience with a CI (minimum 6 months). Attending to participants' device configuration, 60% were unilateral CI users (one CI), 6% were bilateral CI recipients (two CIs) and 34% were bimodal users (one CI and a contralateral HA). Device configuration did not appear to vary systematically across age groups (Figure 4.1). For participants with a unilateral CI, concerning the non-implanted ear, 30% reported being completely deaf, 67% reported having severe-to-profound HL, and 3% reported having mild or moderate HL.

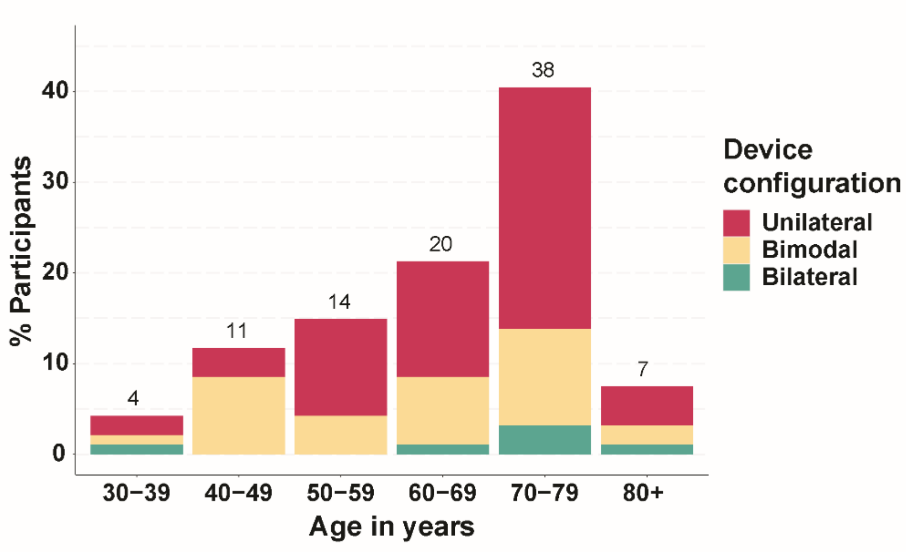


Figure 4.1. *Participants' hearing device configuration by age group. Number of participants in each age group is shown at the top of each bar.*

On average, bimodal listeners had more than ten years' HA experience and 72% reported using their HA more than eight hours a day. Around one-third of bimodal listeners reported making use of a special feature to facilitate coordination between their hearing devices (see Table B. 1 in Appendix B), most commonly, a wireless link between the CI and the HA allowing an audio signal to be transferred between them (e.g., a contralateral routing of signals solution).

Nearly all respondents reported relying mainly on auditory speech for communication, typically with significant support from visual cues including lip reading and facial expressions (Figure 4.2). Around one-half of respondents reported making regular use of text transcriptions to support communication, which included subtitles and speech-to-text transcriptions. Very few participants made use of sign language to communicate with others.

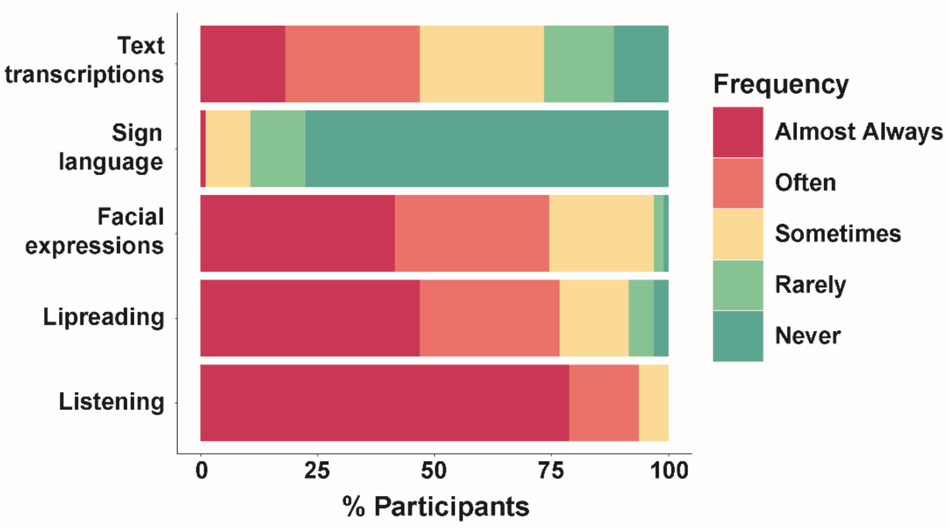


Figure 4.2. *Percentage of participants who rely on different ways of communicating (listening, lip reading, facial expressions, sign language and text transcriptions) in everyday life. Q9: "In everyday life, to what extent do you rely on these ways of communication?".*

4.4.2 Perceived listening difficulties during in-person communication

Nearly all (99%) participants reported having experienced communicating in-person with someone who was wearing a facemask and, separately, whilst maintaining a distance of at least 2 metres. Figure 4.3 shows participants' ratings regarding their current listening experiences under each of these two public-health measures.

Under COVID-19 restrictions, many participants reported experiencing moderate to high levels (scores of 4 or 5 out of 5) of LE (90% and 74% of participants for facemasks and social distancing, respectively), need to ask for repetition (55% and 42%), listening-related anxiety/stress (54% and 45%), and fatigue (58% and 45%). Some, but not all, participants also reported frequently disengaging from listening (31% and 20% for facemasks and social distancing, respectively). Alongside these challenges, many participants reported achieving no better than moderate speech understanding (intelligibility scores ≤3 out of 5) during in-person communication (76% and 48% of participants for facemasks and social distancing, respectively).

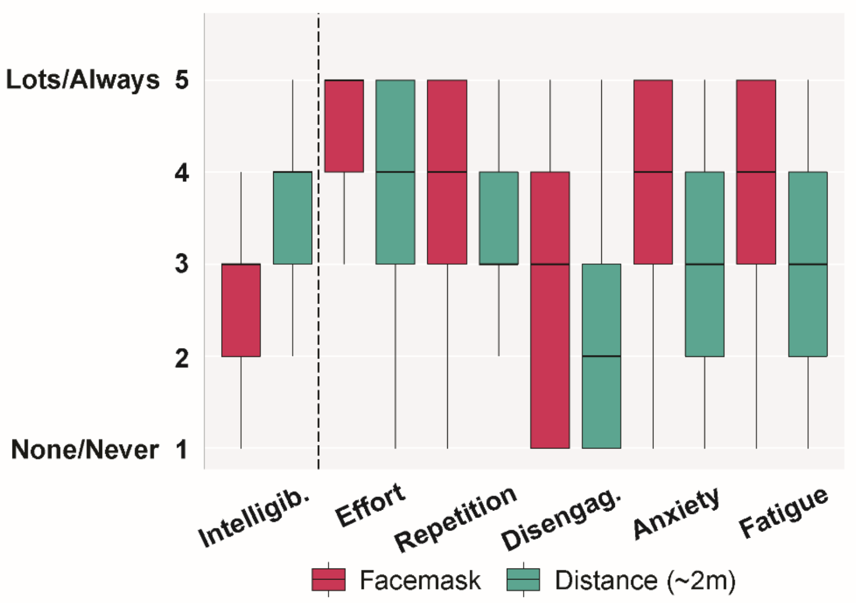


Figure 4.3. *Participants' listening experiences regarding facemasks and social distancing in response to questions Q10.1 and Q13.1. Refer to the main text (under section 4.3. Methods) for the full wording of the questions corresponding to each labelled item on the x-axis. The box represents the inter-quartile range (IQR), with thick lines representing the median.*

The use of facemasks was considered more detrimental to communication than social/physical distancing. Significantly worse ratings ($p < .05$) for all items (listening effort, intelligibility, repetition, disengagement, anxiety/stress, and fatigue) were given in relation to facemasks compared with social distancing (see Table 4.1 for statistical test results).

As well as experiencing significant listening difficulties under COVID-19 public-health measures in place at the time of survey completion, participants reported that their listening experiences had significantly worsened, compare to before the COVID-19 outbreak, specifically because of the widespread use of facemasks and the imposition of social/physical distancing rules (Figure 4.4). This worsening of listening experiences (less perceived intelligibility and more perceived effort, need of repetition, disengagement, anxiety/stress, and fatigue) from before to during the pandemic was statistically significant for all communication items and for both the facemask and social distancing measures (Table 4.1).

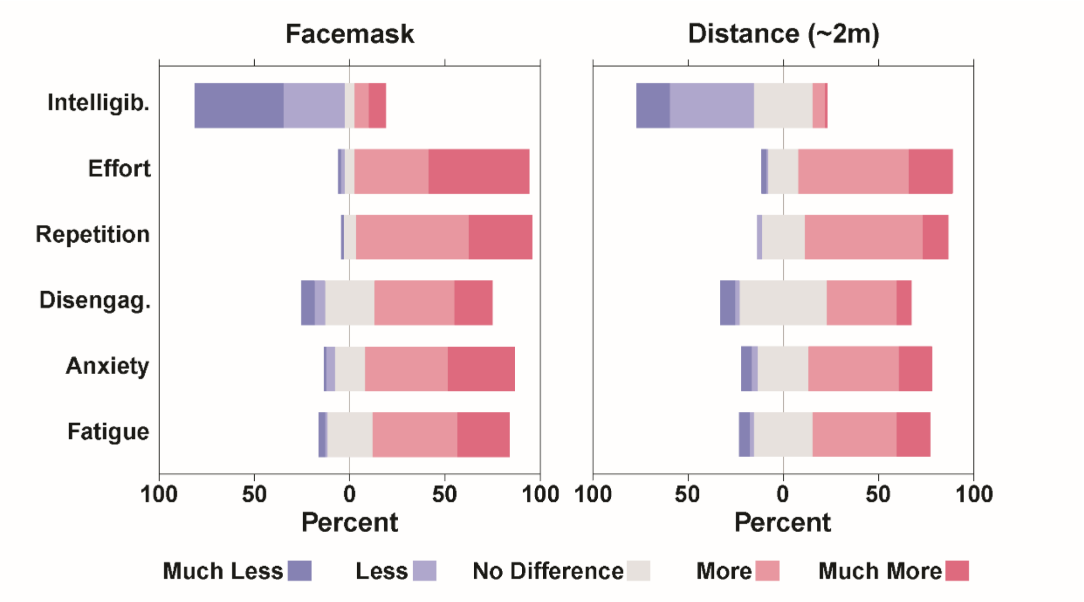


Figure 4.4. *Diverging stacked bar chart showing changes in perceived listening difficulties (before versus after COVID-19 outbreak) due to facemasks and social distancing. Q10.2 and Q13.2: "Considering your listening experiences before and after the COVID-19 outbreak, how much do you think your communications have changed due to the speaker wearing a face mask/due to having to keep 2 metre away from others?". The percentages of participants who perceived "more" or "much more" of each communication item are shown to the right of the zero line in red shades; the percentages of participants who noticed "less" or "much less" of each communication item are shown to the left of the zero line in blue shades; and the percentage of participants who perceived "no difference" are shown centred around the zero line in light grey colour.*

Most participants reported avoiding in-person communication scenarios at least some of the time if either facemask use or social distancing would be required (Figure 4.5). Participants were significantly more likely to avoid scenarios due to the challenges associated with facemask use compared with the challenges associated with maintaining a minimum distance ($p < .0001$).

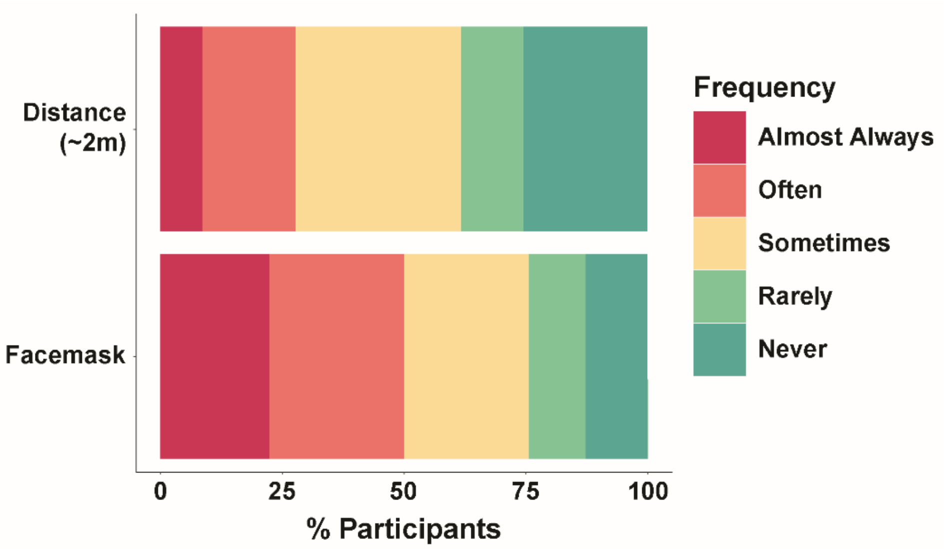


Figure 4.5. *Frequency of communication avoidance due to facemasks and social distancing. Q12 and Q15: "How often do you find yourself avoiding face-to-face communication because of difficulty hearing someone who is wearing a face mask /who is 2 metres apart?"*

Participants identified multiple factors contributing to the listening challenges associated with COVID-19 public-health measures (Figure 4.6). For facemask use, the predominant factors were "Lips not visible" (94% of participants rated it as extremely or very relevant), "no facial expressions" (80%), "muffled sound" (81%) and "quieter voice" (65%). Regarding social/physical distancing, "intrusive background noise", "quieter voice", "difficulty lipreading" and "echoey speech" were rated as extremely or very relevant by 79%, 59%, 59% and 40% of participants, respectively.

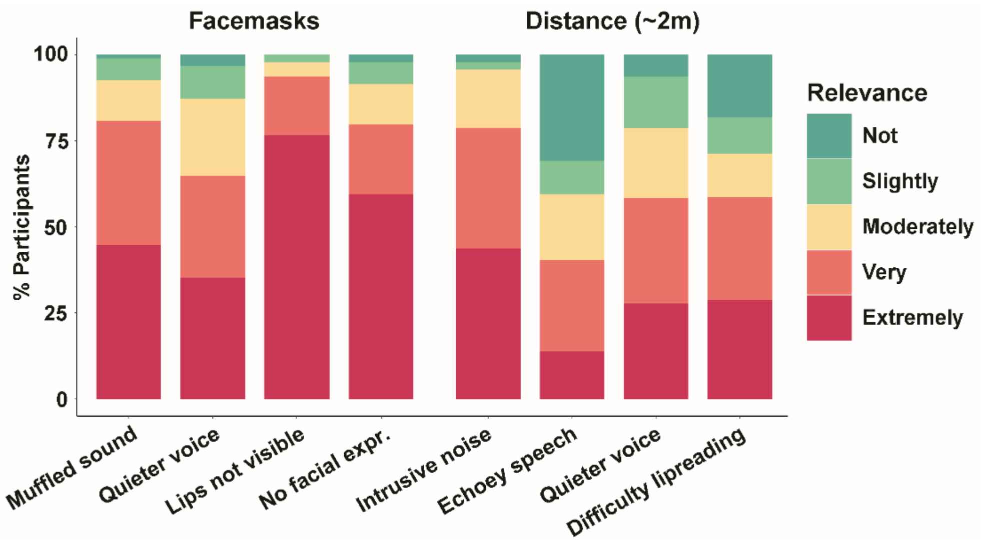


Figure 4.6. *Relevance of listening challenges associated with facemasks and social distancing. Q11 and Q14: "The following is a list of potential challenges associated with listening to someone who is wearing a facemask/ from 2 metre distance. Please rate how relevant they are according to your experience".*

Responses to the open question about in-person communication were provided by 48% of participants. Examples of participants' responses per theme and category can be found in Table B. 2 in Appendix B. Most comments (62%) were related to facemask use ("facemasks" theme). Overall, facemasks were identified as the predominant challenge to successful communication during the COVID-19 pandemic ("major problem" category) due to the inability to lipread and see facial expressions. Some participants commented that facemask use in healthcare settings was especially concerning since it prevents them from understanding important medical information ("in medical settings" category). Some respondents (11% of participants who provided free-text responses) considered that people's collaboration (e.g., temporary removal of facemasks) was needed to overcome the limitations imposed by facemask use ("people collaboration" category). A few participants (7%) stated that avoiding going into places where facemasks would be required had led to loss of confidence, increased feelings of loneliness, and social isolation ("low confidence/ loneliness" category).

Participants expressed varying opinions regarding social/physical distancing (“social distancing” theme). Although some participants (9%) indicated that it is difficult to lipread and understand speech at two-metre distance (“difficult to lip-read” category), others commented (15%) that it was not a problem unless the background noise level was high (“no major impact unless noisy environment” category). Two participants commented that the use of facemasks and social distancing in combination made communication no longer possible (“combination of facemask and social distancing” theme). Plastic shields at counters were also identified (by 4% of participants who provided free-text responses) as a further barrier to successful communication (“Shields” theme).

4.4.3 Perceived listening difficulties during remote communication

Responses to question Q18: “how has the frequency of telephone and video calls changed since the COVID-19 outbreak?” showed that participants reported a significant increase ($p < .0001$) in the frequency of telephone and video calls since the beginning of the pandemic (see “frequency change” under “remote communication” in Table 4.1). Specifically, as shown in Figure 4.7, participants reported that this increased reliance on remote communication, at the time of survey completion, was needed to speak with family and friends (50% often or always), to access essential services (20% often or always), and for work-related reasons (28% often or always). It is worth noting that most participants in our sample were retired, which may account for 37% of participants reporting never having telephone or video calls for work.

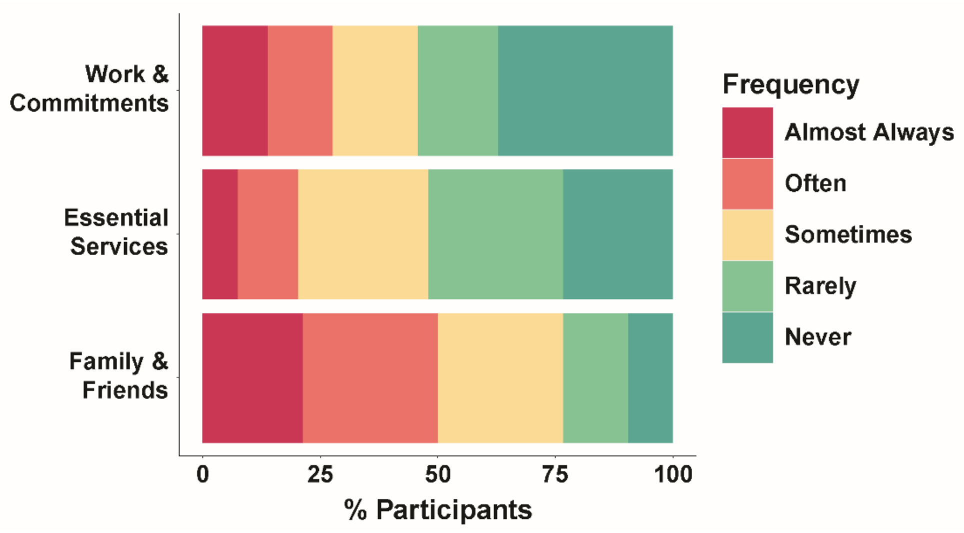


Figure 4.7. *Frequency of telephone and video calls during the COVID-19 pandemic to communicate with family and friends, essential services and for work and other commitments. Q17: "How often do you have telephone or video calls… to communicate with family and friends/… to access essential services (such as health and care consultations, shops, pharmacy, etc.)/…for work or other commitments?"*

Figure 4.8 illustrates participants' listening experiences when having telephone and video calls. Since not all participants made use of remote communication technologies, only 70% of participants (66/94) answered questions about telephone calls while 74% (70/94) answered questions about video calls.

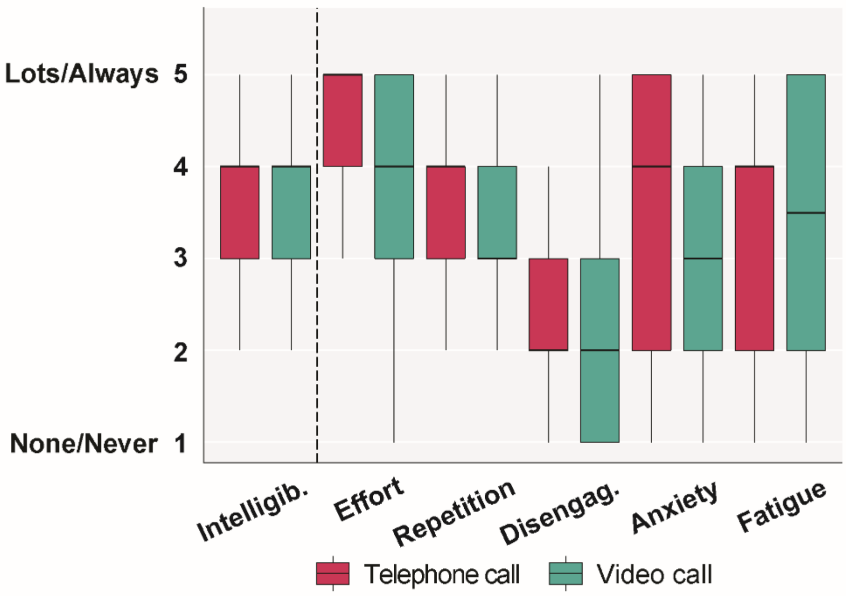


Figure 4.8. *Participants' listening experiences regarding telephone and video calls during Covid-19 pandemic in response to questions Q19.1 and Q21.1. Refer to the main text (under section 4.3. Methods) for the full wording of the questions corresponding to each labelled item on the x-axis.*

Although most participants reported achieving moderate to good levels (scores of 3 or 4 out of 5) of speech intelligibility in both modes of remote communication (85% and 83% of participants for telephone and video calls, respectively), they at the same time reported experiencing relatively high levels (scores of 4 or 5 out of 5) of listening effort (86% and 67% of participants for telephone and video calls, respectively), need to ask for repetition (59% and 37%), listening-related anxiety (58% and 43%), and fatigue (53% and 50%). Despite these difficulties, most respondents reported rarely disengaging (scores ≤2 out of 5) while communicating via telephone (58%) or video call (61%).

Participants' listening experiences were worse when having telephone calls compared to video calls. Pairwise comparisons (Table 4.1 under "remote communication") yielded significant differences in effort ($p < .05$), frequency of repetition ($p < .001$), disengagement ($p < .05$) and anxiety ($p < .01$) between telephone and video calls. No significant differences were found in intelligibility or listening-related fatigue.

As illustrated in Figure 4.9, most participants avoided both modes of remote communication at least some of the time. However, participants were more likely ($p < .0001$) to avoid telephone calls than video calls.

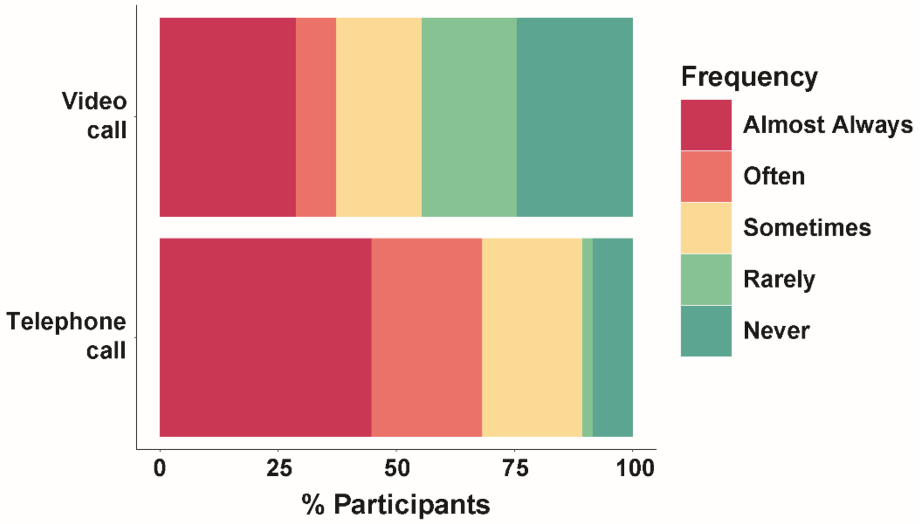


Figure 4.9. *Frequency of telephone and video calls avoidance during the COVID-19 pandemic. Q20 and Q22: "How often do you find yourself avoiding telephone calls/ video calls because of difficulty understanding what is being said?"*

All potential challenges associated with remote communication proposed in the survey were considered very or extremely relevant by more than 50% of participants (Figure 4.10). The primary problems associated with telephone calls were related to the speaker's voice ("unfamiliar voice or accent") and pace ("fast speech"). Also relevant were "poor quality line", background noise in the participant's environment ("noisy environment") and "volume too low". With respect to video calls, once more "fast speech pace" was the most relevant challenge, followed by competing speech in multitalker conversations ("too many people speaking at the same time") and connection problems ("audio or video cutting in and out"). Other relevant problems were background noise ("noisy environment"), "volume too low", unclear who was speaking during group conversations ("who speaks?"), "audio and video out of sync", "poor or lack of transcriptions", "no access to the speaker's video camera", and "unfamiliar voice or accent".

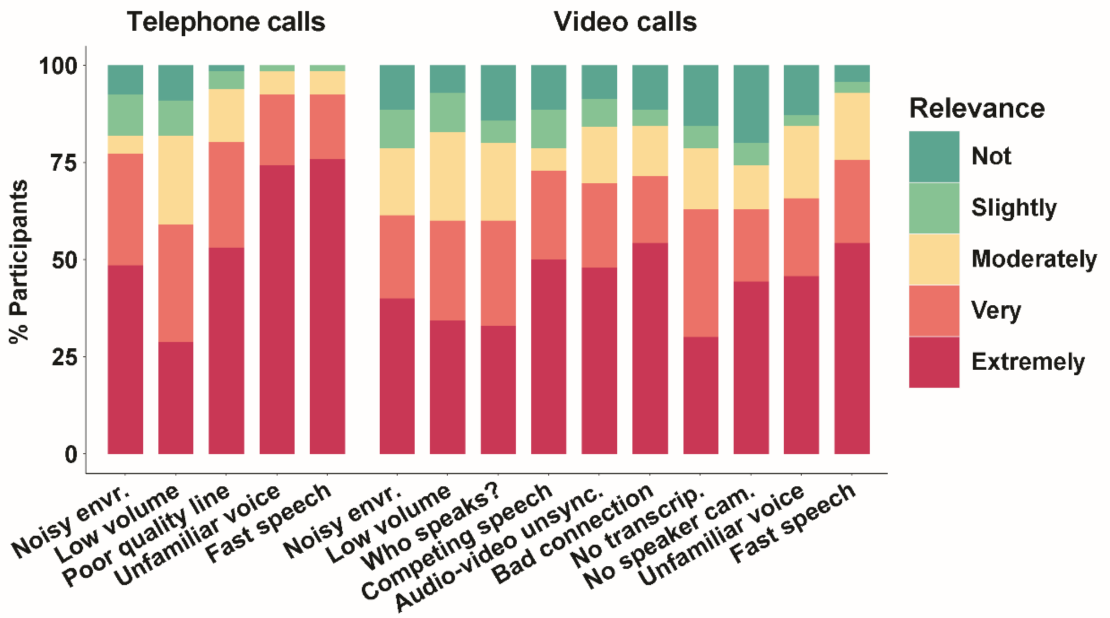


Figure 4.10. *Relevance of potential challenges associated with telephone and video calls. Q19.2 and Q21.2: "The following is a list of potential challenges associated with telephone conversations/ video calls or conferences. Please rate how relevant they are according to your experience".*

Table 4.1. *Pairwise and single-sample comparisons using Wilcoxon signed-rank test with continuity correction. N represents the number of complete observations included in each test. P values are adjusted for multiple comparisons. Significance code: ns (not significant), * (p < .05), ** (p < .01), *** (p < .001), **** (p < .0001).*

	Variable	**N**	**Comparison**	**Z**	**P value adjusted (two-sided)**	**P value Signif.**
In-person	Effort (EF)	92	Facemask vs. ~2m distance	926	0.000019	****
	Intelligibility (INT)	92	Facemask vs. ~2m distance	184.5	0.000002	****
	Repetition	92	Facemask vs. ~2m distance	897	0.018	*
	Disengagement (DISG)	92	Facemask vs. ~2m distance	824.5	0.019	*
	Anxiety	92	Facemask vs. ~2m distance	1053	0.018	*
	Fatigue	92	Facemask vs. ~2m distance	956.5	0.012	*
	EF Change (Facemask)	93	Communication change due to facemask vs. “No difference”	3813.5	< .000001	****
	EF Change (~2m distance)	93	Communication change due to social distance vs. “No difference”	2919	< .000001	****
	INT Change (Facemask)	93	Communication change due to facemask vs. “No difference”	637	< .000001	****
	INT Change (~2m distance)	93	Communication change due to social distance vs. “No difference”	200	< .000001	****
	Repetition Change (Facemask)	91	Communication change due to facemask vs. “No difference”	3585	< .000001	****
	Repetition Change (~2m distance)	92	Communication change due to social distance vs. “No difference”	2496	< .000001	****
	DISG Change (Facemask)	91	Communication change due to facemask vs. “No difference”	1835	0.000074	****
	DISG Change (~2m distance)	93	Communication change due to social distance vs. “No difference”	933.5	0.018	*
	Anxiety Change (Facemask)	92	Communication change due to facemask vs. “No difference”	2852	< .000001	****
	Anxiety Change (~2m distance)	93	Communication change due to social distance vs. “No difference”	1984	0.000003	****
	Fatigue Change (Facemask)	92	Communication change due to facemask vs. “No difference”	2294	< .000001	****
	Fatigue Change (~2m distance)	93	Communication change due to social distance vs. “No difference”	1766	0.000006	****
	Avoidance	94	Facemask vs. ~2m distance	1341.5	0.000067	****
Remote	Effort (EF)	54	Telephone vs. Video call	331.5	0.018	*
	Intelligibility (INT)	54	Telephone vs. Video call	149	0.321	ns
	Repetition	54	Telephone vs. Video call	507	0.000186	***
	Disengagement	54	Telephone vs. Video call	334.5	0.026	*
	Anxiety	54	Telephone vs. Video call	490	0.001	**
	Fatigue	54	Telephone vs. Video call	299.5	0.321	ns
	Frequency Change	94	Telephone & Video call frequency change vs. “No difference”	1666	< .000001	****
	Avoidance	94	Telephone vs. Video call	1576	0.000002	****

4.4.4 Factor analysis

EFA was conducted to explore whether the six communication items that we used to probe participants' listening experiences each provided unique information, or whether they tapped into one or more common underlying constructs.

Separate EFA analyses were conducted for each communication scenario: participants' experience communicating with someone wearing a facemask (Q10.1), at two-metre distance (Q13.1), having telephone calls (Q19.1), and video calls (Q21.1). In all cases, the solutions provided by the function 'nScree' (Kaiser rule, parallel analysis, acceleration factor, and optimal coordinates index) indicated a clear one-factor structure. The one-factor model explained 52%, 64%, 55% and 53% of the total variance for each scenario, respectively. The models showed a consistent pattern of factor loadings as revealed by a Tucker's congruence coefficient (φ) ≥ 0.99 across all pairwise comparisons (Lorenzo-Seva and ten Berge 2006).

Figure 4.11 shows the resulting factor loadings for each communication scenario with 95% bootstrap confidence intervals. The factor loading of a variable quantifies the extent to which the variable is related with the underlying factor. A factor loading of more than 0.5 usually indicates good correlation between the variable and the factor. All items had broadly similar loadings (> 0.5) on the principal underlying factor across the four scenarios, with the confidence intervals generally overlapping. A possible exception was the "anxiety/stress" item, which showed a consistently high loading across the four scenarios, with relatively small confidence intervals compared to the other items.

Overall, the results of the EFAs suggest that, across multiple communication scenarios, the six communication items all tapped into a single underlying construct that reflected both immediate listening difficulties (reduced intelligibility and increased listening effort) as well as short-term (need for repetition and risk of disengagement) and longer-term consequences (anxiety and fatigue).

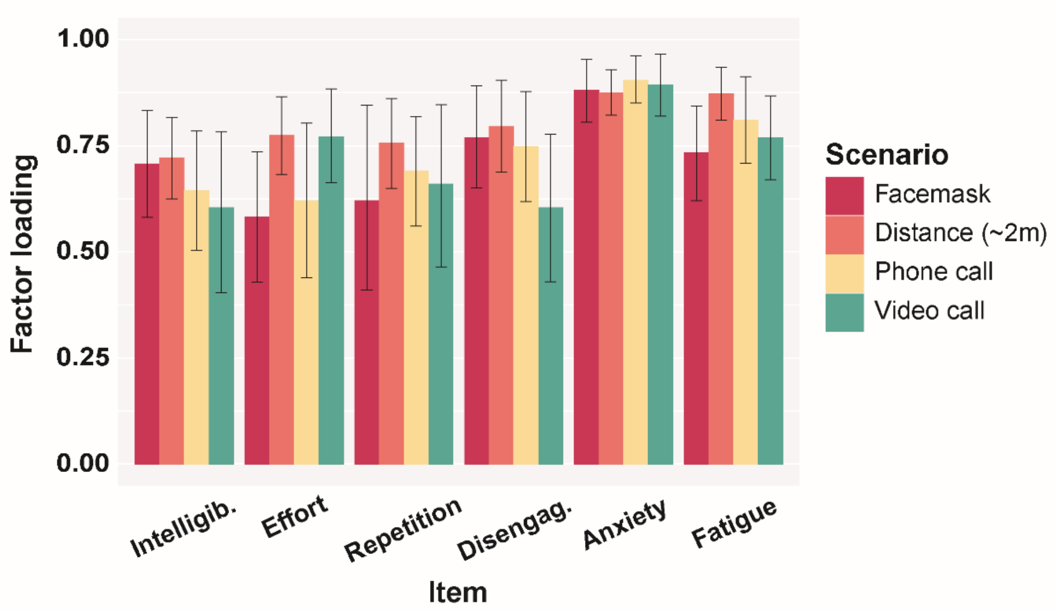


Figure 4.11. *Factor loadings resulting from exploratory factor analysis across four communication scenarios (facemask, social distance, phone and video calls). Error bars represent 95% bootstrap confidence intervals.*

4.4.5 Solutions to minimise the impact

Most participants considered, as illustrated in Figure 4.12, that the solutions proposed in the survey could help them greatly to improve their everyday life communication.

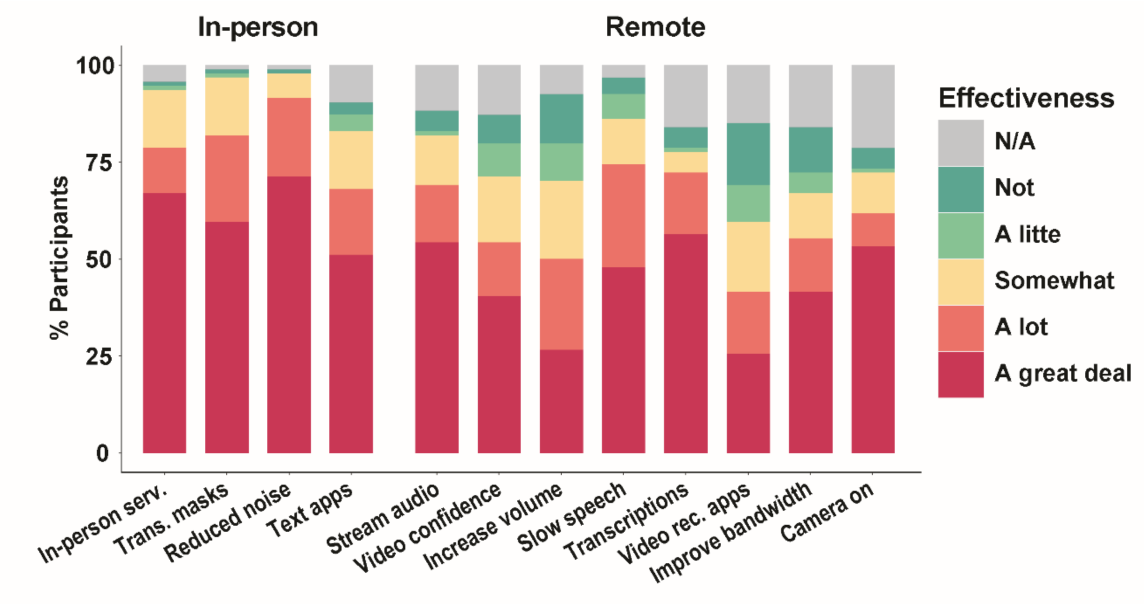


Figure 4.12. *Effectiveness of potential solutions to improve in-person and remote communication. Q23: "To what extent do you think these solutions could help to improve your everyday life communications? If any of these solutions doesn't apply to you, please check the "Not applicable" option". Proposed solutions: "In-person serv." (more access to face-to-face services); "Trans. mask" (transparent face masks); "Reduced noise" (reduced background noise in public places); "Text apps" (speech-to-text apps); "Stream audio" (stream sounds from phone and video call directly to hearing devices); "Video confidence" (improved confidence to use video calling); "Increase volume" (increased volume in phone and video calls); "Slow speech" (slower speech pace); "Transcriptions" (real-time transcriptions during video calls); "Video rec. apps" (video call recording apps); "Improve bandwidth" (improved bandwidth during video calls); and "Camera on" (speaker's camera turned on during video calls).*

As reported in Table 4.2, the most highly rated solutions to improve in-person conversations were the reduction of background noise in public places (91% of participants rated it as highly effective) and the use of transparent face masks (82%). Also relevant were having more access to in-person services (79%), and the use of speech-to-text apps (68%).

For remote conversations, the most highly rated solutions were: the speaker talking at a slower speech pace (75%), real-time transcriptions during video conferences (72%), streaming sounds from phone and video call directly to their hearing devices (69%), and making sure that the speaker's camera is turned on during video calls (62%). Other solutions rated as highly effective by at least half of participants were improved bandwidth during video calls, improved confidence using video conferencing, and increased volume in phone and video calls.

Table 4.2*. Potential solutions to improve in-person and remote communication and percentage of participants rating each solution as being expected to be highly effective, slightly effective, or not effective/ not applicable.*

Type	**Potential Solutions**	**% of Participants**		
		Highly effective *(A lot/ A great deal)*	**Slightly effective** *(A little/ somewhat)*	**Not effective/ Not applicable**
IN-PERSON	More access to face-to-face services	78.7	16	5.3
	The use of transparent face masks	81.9	16	2.1
	Reduced background noise in public places	91.5	6.4	2.1
	Use of speech-to-text (live transcription) apps	68.1	19.1	12.8
REMOTE	Stream sounds from phone and video call directly to hearing devices	69.1	13.8	17
	Improved confidence to use video calling	54.3	25.5	20.2
	Increased volume in phone and video calls	50	29.8	20.2
	Slower speech pace	74.5	18.1	7.4
	Real-time transcriptions during video calls	72.3	6.4	21.3
	Video call recording apps (allowing re-watching of the call afterwards)	41.5	27.7	30.9
	Improved bandwidth during video calls (less cutting out of audio or video)	55.3	17	27.7
	Making sure that the person speaking always has their camera turned on during video calls	61.7	11.7	26.6

Around one half of participants provided additional free-text comments about solutions that they were already using to improve daily-life conversations (see Table B. 3 in Appendix B for description of themes, categories and statement examples). With respect to remote communication, the use of text transcriptions during video calls was commonly mentioned (31%) as a solution that participants were already employing ("text transcriptions/live subtitles" theme). Participants accessed live subtitles from some video call platforms or from external transcription services and mobile phone apps (e.g., Live Transcribe) that allow capturing speech-to-text in real time ("text transcriptions apps" category). One downside mentioned about live captions was that they are sometimes inaccurate ("not always accurate" category). Another popular solution (14%) used by participants during telephone and video calls was streaming sounds directly to their hearing devices ("stream sounds to hearing devices” theme). This solution however is not as effective when people speak at the same time (competing speech) during group conversations (“overlapping talk” category). One participant also found the use of over-the-ear headphones helpful to improve audio quality during remote conversations (“other solutions” theme).

For those who completely avoid telephone and video calls (14%) (“telephone and video call avoidance” theme), the preferred communication method was the use of text-messaging apps or other services such as RelayUK that offers an intermediate assistant who can speak on the CI user’s behalf (“use of text messaging and other services” category).

Regarding in-person communication, the use of an external mini microphone connected to the hearing device was mentioned (8% of comments) as a good option to cope with the difficulties of social/physical distancing (“External microphone” theme). However, one responder expressed concerns about the potential risk of COVID-19 transmission when the microphone is handled by multiple people, which led that individual to stop using an external microphone. Asking people to briefly remove their facemasks (“ask people collaboration” theme), wearing a badge saying “I am deaf” (“other solutions” theme), and encouraging the use of transparent facemasks and face visors (“transparent masks/face visors” theme) were other strategies that individual participants (18%) had employed during the pandemic.

4.5 DISCUSSION

Ninety-four English-speaking adult CI users completed an online survey asking about perceived listening difficulties during in-person and remote communication under public-health measures introduced to control the spread of COVID-19. Respondents also gave their opinions regarding suggested strategies and technological solutions that could help CI users to overcome some of the listening challenges associated with social distancing measures and online communication.

Perceived listening difficulties during in-person and remote communication: a single underlying construct

Across multiple communication scenarios, participants reported experiencing a diverse array of listening difficulties during the COVID-19 pandemic, including limited intelligibility, effortful listening, need to ask for repetition of a message, disengagement, and feelings of listening-related stress/anxiety, and fatigue. Statistical analysis confirmed that participants considered their listening difficulties during in-person communication to have become significantly worse (for all communication items queried) compared to pre-pandemic times because of the public-health measures that had been introduced (facemasks and social/physical distancing). These issues were sufficiently troublesome that most participants reported actively avoiding certain communication scenarios.

Although there were some differences across communication scenarios, there were also commonalities. Ratings of LE were consistently high for both in-person and remote communication. This finding was observed regardless of the level of speech understanding achieved, which was higher in the remote communication scenarios. This supports the notion that listening through a CI can be cognitively effortful, even when intelligibility remains high (Pals et al. 2020, 2013; Winn et al. 2015; Winn and Teece 2021)

Whilst participants did report actively avoiding challenging communication scenarios during the pandemic, ratings for listening disengagement were consistently lowest amongst the six items. This suggests that, although most participants were avoiding some situations altogether due to the listening challenges involved, once actually engaged in an interaction, participants generally persevered with trying to keep up communication. This could be explained by motivational factors (Eckert et al. 2016; Herrmann and Johnsrude 2020), considering that the need to communicate with others during the pandemic (and the benefits that communication can bring) may have surpassed the cognitive cost of doing so.

Despite evidence of a diverse array of perceived listening difficulties experienced by adult CI users during the COVID-19 pandemic, EFA in all cases suggested that the data were best explained by a single underlying factor (interpreted by the authors as representing "overall listening difficulty"). Thus, rather than representing distinct and independent dimensions, our data suggest a strong interconnection between immediate listening challenges (reduced intelligibility, high effort), short-term implications (need to ask for repetition, risk of disengaging), and longer-term consequences (stress/anxiety and fatigue). It is noteworthy that, across the different communication scenarios, the stress/anxiety item received the highest and most consistent factor loading scores. One cannot rule out the possibility that the negative experiences reported by participants in the survey may have been in part influenced by general feelings of stress and anxiety associated with living through the pandemic.

Changes in communication during the pandemic

The ways in which people communicate changed dramatically following the outbreak of the COVID-19 pandemic, as governments and institutions adopted widespread public-health measures to limit the spread of the virus. Results of this survey corroborate that the use of facemasks and the imposition of social/physical distancing rules posed additional listening challenges to CI users, which has led to them at times actively avoiding certain communication scenarios. Moreover, much like for the wider population, the adult CI users who completed this study reported a significant increase in the frequency of telephone and video calls. Concerningly, many respondents reported regularly avoiding remote communications due to the listening challenges involved. The World Health Organization (WHO) issued guidance during the COVID-19 pandemic specifically advising people to stay connected to friends, family and community members via remote communication to mitigate the psychological effects associated with sustained periods of isolation (WHO 2020b). Avoiding remote communication completely may expose an individual to higher risk of suffering psychological harm due to social isolation (Razai et al. 2020). Our results add to a growing body of evidence that the pandemic had a negative and far-reaching impact on communication, especially amongst people with HL, which contributed to heightened feelings of stress, anxiety, and fatigue (Ideas for Ears Ltd 2020; Saunders et al. 2020; Tagupa 2020; Tavanai et al. 2021).

Nonetheless, as previous research has highlighted (Dunn et al. 2020; Naylor et al. 2020), not all changes associated with the COVID-19 pandemic have been negative. Dunn et al. found that social distancing measures promote people spending more time at home and in generally quieter environments, where more favourable SNRs are present. Dunn et al. concluded that CI users' listening experiences under such circumstances were more positive, being associated with better speech understanding and less LE. Overall, feelings of social isolation and anxiety due to HL were reduced during the COVID-19 pandemic, compared to before its outbreak, since due to the lack of group interactions there were fewer occasions where participants felt left out of conversations because of their HL. Similar results were found in Naylor et al.'s study: participants with greater HL showed substantial relief at avoiding social gatherings. Nonetheless, the lack of social interactions during the pandemic could also bring increased feelings of loneliness as participants in Dunn et al.'s study reported. These findings highlight the importance of taking a holistic view of CI users' listening experiences, which involves not just the additional burdens imposed by COVID-19 related public-health measures (the focus of the present study), but also possible positive effects associated with individual changes in auditory ecology. The survey administered in the present study captured limited information about wider changes in auditory ecology, beyond the specific scenarios that participants were questioned about.

The importance of visual cues

The results of the present survey evidence the importance of visual cues to CI users as an aid to speech understanding. Most participants reported relying significantly on visual cues, such as lipreading and facial expressions, to support their everyday communication. Therefore, it is unsurprising that the use of facemasks was considered to have the greatest detrimental impact on in-person communication among the COVID-19 measures considered. These results are in line with Naylor et al.'s (2020) study, which found that participants with HL reported better communication performance under social distancing conditions compared to facemask use. Communication difficulties associated with the obscuring of the speaker’s mouth and lower part of the face motivated some participants in the present study to avoid scenarios where the use of facemasks would be mandatory. According to participants’ free-text comments, the use of facemasks was particularly concerning in medical settings since CI users feared mishearing or misinterpreting important information that could affect their health. Similar results were found by Saunders et al. (2020) who reported face coverings to have a greater negative impact on communication in medical situations (e.g., doctor’s appointments, pharmacist and hospital visits) compared to other social interactions (family/friends, shop assistants, at work). Transparent facemasks and clear face visors were identified by participants in the present study as an efficient solution to overcome this issue. Indeed, it is known from previous research (Atcherson et al. 2017) that the use of transparent facemasks significantly improves the level of speech understanding achieved by participants with severe-to-profound HL, even in the presence of background noise.

Similarly, the absence of visual cues meant that participants in the present survey reported telephone calls as being more challenging (and hence more frequently avoided) than video calls. Participants did, however, emphasize the importance of the speaker having their video camera turned on for the benefits of video calling to be realised. Live captions during video calls were considered another important feature that provides visual cues to support communication. Indeed, many participants highlighted that live speech-to-text transcriptions should be made available across all video-calling platforms. A similar observation was made in Chodosh et al.’s study (Chodosh et al. 2020), which identified access to free online captions as a priority for innovation and inclusive communication.

Recommendations to make in-person and remote communication easier for people with a CI

As discussed in the preceding sections, the results of the present survey highlight that solutions that offer improved access to visual cues should be adopted wherever practical. For in-person communication, this could involve the use of a transparent facemask, or the use of a clear face visor in place of a facemask. According to the WHO (WHO 2020a), while face shields and visors provide inferior protection against COVID-19 transmission compared to masks, they are considered valid alternative solutions for the deaf and hard of hearing community. For visors to provide a good level of protection in short exposure situations, they should cover the entire face, above the eyes to below the chin and wrap around from ear to ear (Wendling et al. 2021). Nonetheless, social distance must be maintained in combination to face shield use to provide additional protection against smaller particles that can remain airborne for longer periods of time (Lindsley et al. 2014). For remote communication, video calls are to be preferred over telephone calls wherever possible, with care taken to ensure that cameras are turned on allowing clear visibility of the face. In addition to ensuring access to visual speech cues, both in-person and remote communication can potentially be further supported using software or mobile app solutions that offer live speech-to-text transcriptions.

When communicating in person, our results suggest that, where such an arrangement is considered safe, most adult CI users would feel able to communicate more effectively if standing further away from someone who was *not* wearing a facemask, compared to if they were standing closer to someone who *was* wearing a facemask. However, for communication to succeed at greater distances, the quality of the acoustic signal arriving at the listener's ears must be adequately preserved. As mentioned in the introduction, with increased distance between conversational partners comes a reduction in SNR and a reduction in the ratio of direct to reverberant sound. As CI users are particularly susceptible to the adverse effects of both background noise and reverberation (Badajoz-Davila et al. 2020; Hazrati and Loizou 2012; Kressner et al. 2018), this can cause significant problems. Background noise can be reduced by directly controlling sound from any unwanted sources, e.g., by turning background music down or off, and by ensuring that air conditioning units are operating quietly and efficiently. Both background noise and reverberation can be controlled effectively through the introduction of simple acoustic treatment in the form of sound absorbing materials (e.g., soft furnishings) or dedicated acoustic wall or ceiling panels. Such measures will be especially effective in rooms that otherwise feature mostly hard, reflective surfaces and which are likely to be excessively reverberant to begin with. However, such solutions cannot be controlled or adopted by CI users first-hand, but only by those responsible for the upkeep and operation of public venues.

A solution that can be directly implemented by CI users to improve the quality of the acoustic signal during in-person communication is the use of an external mini-microphone that can be wirelessly connected to the CI user's hearing device(s). These systems largely overcome the deleterious effects of background noise and reverberation by picking up the target speech signal close to its source and then transmitting it wirelessly (with minimal degradation) to the listener's ears. This is a powerful technological solution, but one that might not always be practical in public locations given that its use typically requires cooperation from the communication partner. Ensuring adequate sanitisation of the equipment to avoid any risk of surface-borne transmission of the virus is often recommended, as long as pandemic conditions continue to prevail.

Maximising the quality of the acoustic signal is also important when it comes to remote communication. Some video calling systems now provide built-in noise reduction to ensure that speech signals are picked up as cleanly as possible at source. Similarly, both passive (e.g., turning off any unnecessary sources of background noise) and active (e.g., use of noise-cancelling headphones) actions to reduce noise in the CI user's physical environment may be helpful. Similar benefits may be derived from streaming the sound directly from the computer/tablet/phone to the hearing device(s) using a wireless or wired (direct input) connection.

Finally, the importance of simple behavioural adjustments should not be undervalued when it comes to facilitating effective verbal communication. Participants identified a slower speech pace as being beneficial for both in-person and remote communication. This is consistent with prior research which evidenced that slowed speaking rate provides release from LE in CI users as measured by behavioural and pupillometry techniques (Winn and Teece 2020).

4.5.1 Limitations

Recruitment into the present study was conducted online, via email and social media. This may have introduced selection bias, since participants volunteering to complete an online survey may not be representative of the wider population of CI users. Likewise, our sample was unbalanced in terms of gender, nationality, and age, most participants being females from the UK, in their seventies, and retired. This could have influenced the results associated with remote communication since older people may be more likely to avoid these technologies or use them to communicate with family and friends rather than for work. However, the age distribution of our participants may not be entirely unrepresentative given that many adult recipients of a CI are aged 60-69 at the time of implantation according to the UK surgical registration data (Raine 2014).

Another limitation of the study is the lack of pre-pandemic baseline data, which makes assessment of pre- versus peri- COVID-19 listening experiences subject to possible recall bias. This is a common limitation of retrospective questionnaires. Similarly, the lack of data from a control group, for instance, people with NH or a lesser degree of HL, means that it is not possible to say how specific our findings are to the CI-using population. It is possible that other groups, perhaps even everyone, experienced increased listening difficulties because of the public-health measures introduced to combat the spread of COVID-19.

While EFA suggested that the six communication items all loaded on to a single underlying factor, it is possible that intercorrelation amongst items was elevated by the fact that participants answered all questions at the same time, in a fixed order. A finer distinction between different domains of perceived listening difficulty may have been obtained using an alternative methodology (e.g., ecological momentary assessment).

A further limitation is the assessment of participants’ communication experiences of facemask use and social/physical distancing separately, which may not have adequately reflected the everyday reality that these measures tended to be used in conjunction. Open (free-text) questions, however, were able to collect participants’ opinions and experiences in that regard, with some people noting that the combination of facemasks plus distancing was especially problematic.

Finally, although all questions were explicitly hearing focused, other psychological factors prevalent during the COVID-19 pandemic such as general health anxiety and loneliness may have influenced participants' responses.

4.6 CONCLUSION

Adult CI users' in-person listening experiences in some common everyday scenarios worsened significantly during the COVID-19 pandemic due to the widespread use of facemasks and the imposition of physical distancing rules to control the spread of the virus. Participants reported experiencing an array of listening difficulties, including reduced speech intelligibility and increased LE, which resulted in many CI users actively avoiding certain communication scenarios at least some of the time. CI users also experienced similar listening difficulties during remote communication, though the frequency with which they held telephone and video calls increased significantly during the pandemic. The results suggest ways in which everyday communication might be made easier for people with CIs, both during the pandemic and beyond. The importance of visual cues was evident for both in-person and remote communication. Solutions that offer improved access to visual cues (e.g., transparent instead of opaque face coverings, video calls instead of telephone calls, live speech-to-text subtitling) should therefore be adopted whenever possible. The results also highlighted the potential importance of relatively simple behavioural (e.g., slowed speaking rate) and environmental (e.g., control of background noise and reverberation in public places) modifications that could help to relieve the cognitive burden of everyday listening with a CI.

CHAPTER 5. CI USERS' PREFERENCES AND LISTENING EFFICIENCY DURING ONLINE VIDEO CALLS

5.1 CHAPTER OVERVIEW

During the COVID-19 pandemic, global society has experienced a rapid replacement of in-person interactions by remote communication. While online communication tools hold potential to improve accessibility, previous studies have suggested that the increased reliance on remote communication triggered by the pandemic posed additional communication challenges to people with HI, including CI users. To investigate the cognitive demands of online communication, we designed an online test that explored CI users' preferences and listening efficiency when attending to video calls under three presentation modes: audio (audio only), video (audio-visual), and captions (video plus captions). Fifty adult CI users and fifty approximately age-matched NH controls participated in the test. They were asked to provide their subjective preferences about which presentation mode they would have chosen in real life when viewing examples of pre-recorded online conversations. Subsequently, participants performed a behavioural test of speech recognition under the same three presentation modes, in which accuracy and response time were recorded. To calculate listening efficiency a joint analysis of accuracy and response time was performed using a hierarchical LBA model. Results showed that while NH controls opted for the traditional video mode, most CI users preferred the caption presentation mode since this feature allowed them to follow the conversation easily. These preferences were confirmed by the behavioural performance. Indeed, the presence of visual cues improved considerably CI users' listening efficiency. Results revealed that CI listeners benefited significantly from the addition of video (i.e., being able to see the talker's face) and benefited further from the addition of captions. Nonetheless, even with captions, CI users' listening efficiency remained significantly below that achieved by NH controls in all presentation modes. This suggests that online communication may be more cognitively demanding for CI users than for their NH peers. Participants in the NH group achieved similar performance regardless of presentation mode. These findings provide objective evidence of the listening advantages that captions provide to the CI community during online video calls. We therefore considered that online communication platforms should make them available and adjustable according to individual need and preference.

5.2 INTRODUCTION

The COVID-19 pandemic has affected greatly our daily life communication, which has transitioned from in-person to remote interactions. The reliance on remote communication increased as social encounters became restricted during various stages of the pandemic. As a result, a so-called “new normality” emerged where our social and working life, as well as the access to essential services, needed to take place remotely. Such transition towards online delivered services reached almost all sectors of society, including healthcare services (Bokolo 2020; Reay et al. 2020; White et al. 2021), education (Hyseni Duraku and Hoxha 2020), retails (Bhatti et al. 2020), and remote working (Kniffin et al. 2021; Phillips 2020; Wang et al. 2021). Unavoidably, it led to internet-based technologies becoming the predominant means of communication.

The use of technology during online communication is sometimes challenging since technical difficulties may arise such as bandwidth-related limitations in speech quality, stuttering video streams, and potential lags between audio and video streams that can distort communication and create cognitively demanding listening conditions. Such disruptions could impede effective communication for everyone, but especially for people with HL. In fact, remote communication is considered an important challenge for the HL community. Previous studies showed that the increased reliance on remote communication during the pandemic imposed an additional burden on people with HI (Ideas for Ears Ltd 2020; Naylor et al. 2020; Saunders and Oliver 2022; Tavanai et al. 2021).

Among the HL community, CI users might be greatly affected since they have a higher degree (i.e., severe-to-profound) of HL. A recent study showed that, among adults with HL, CI users experienced greater challenges in communication and healthcare access during the COVID-19 pandemic (Wilson et al. 2021). Naylor et al.'s study also showed that participants with greater HL generally reported inferior performance during video calls compared with in-person conversations. Likewise, the online survey (Perea Pérez et al. 2022) described in CHAPTER 4, showed that CI users reported high levels of listening effort, among other listening difficulties, during telephone and video calls, which led them to avoid both modes of remote communication. Overcoming such difficulties was important during the Covid-19 pandemic to stay connected to others and mitigate the psychological effects associated with sustained periods of isolation (WHO 2020b). Nonetheless, effective virtual communication will continue to be relevant in post-COVID societies where online communication is likely to remain prevalent (Kane et al. 2021; Masalimova et al. 2021).

Several associations aware of the difficulties that people with HL faced during online communication, published guidelines providing advice for effective communication during virtual meetings (ASHA 2020; Maru et al. 2021; NDC 2020; Reed, Ferrante, and Oh 2020; RNID 2022; UCL 2020). These recommendations covered different aspects of the communication experience such as the speaker's behaviour (e.g., facing the camera, speaking slowly and in turns, rephrasing rather than repeating information), environmental factors (e.g., reduced background, good lighting), and technological solutions (e.g., video camera turned on, live transcriptions, headphones, microphone, meeting recordings, streaming sounds to hearing devices).

Regarding the latter, and considering the rapid development of technology, it is not surprising that the use of certain video call features is usually recommended to improve accessibility. For instance, the use of a video camera during virtual meetings is usually advised since having visual access to the speakers' face support speech comprehension (allowing lipreading and seeing facial expressions). Previous research have already proved that CI listeners rely on both auditory and visual information to optimise communication performance (Moberly et al. 2020). These benefits are not limited to people with HL, they are also perceived by NH individuals (Bernstein, Auer, and Takayanagi 2004). Indeed, the McGurk effect is an illusion that demonstrates the critical importance of interactions between hearing and vision in speech perception.

It showed that when incongruent audiovisual stimuli are presented (e.g. the syllables 'ba' are spoken over the lip movements of 'ga') the brain perceives a third phoneme (e.g. 'da') which is different from the one spoken or mouthed, regardless of the listener's hearing status.

The advantages of live transcriptions however are not so evident. Although the availability of captions offers clear advantages in terms of inclusivity, there is no consensus yet as to whether they enhance comprehension. It would certainly depend on individuals' degree of HL and lip reading ability. While the use of captions is optional for listeners with some residual hearing and good lip reading ability, they may be essential for those with profound HL who cannot lipread. Previous research found that captioning helps in improving the understanding of televised content in both deaf students and HA users (Gernsbacher 2015; Jelinek Lewis and Jackson 2001; Sharma and Raghunath Rao 2017). However, other research showed that captions do not significantly change the overall level of information assimilation of individuals with HL (Gulliver and Ghinea 2003). Instead, Gulliver and Ghinea's study concluded that captions caused a shift in attention from video information to caption information, which provided hearing-impaired participants with a greater level of context of the video. Such a shift in attention is not surprising given that the assimilation of audio, visual and textual information concurrently is difficult (Ghinea and Thomas 1998). A study investigating the eye movement of six NH individuals while watching video segments with and without captions confirmed this (Jensema et al. 2000). The authors found that the addition of captions resulted in major changes in eye movement patterns, with the viewing process becoming primarily a reading process. Nonetheless, they noticed individual differences, so that people who were accustomed to lipreading spent more time looking at the actors' lips.

Considering this, it would be expected that when captions are available, CI users would keep shifting attention between visual and textual information so they can both read the subtitles and lipread. Therefore, it is unclear whether the benefits of having extra information to support speech comprehension would come at the cost of increased cognitive load. It is plausible that reading live (and oftentimes imperfect) subtitles while also trying to attend to the speaker's voice and lip movements could be cognitively demanding.

Moreover, videoconferencing is already considered a tiring task by many. Indeed, the term "Zoom fatigue" or "videoconference fatigue" was coined in March 2020 to describe the feelings of being overly drained after a period of meeting over a videoconference tool (Bennett et al. 2021). One of the explanations proposed to justify this phenomenon was that during video calls more attention is paid to non-verbal cues like facial expressions, the tone and pitch of the voice, and body language, which consume a lot of energy. CI users being already accustomed to pay a great deal of attention to non-verbal cues may not notice any extra level of fatigue (Aspinall 2021). However, there are other challenges that could contribute to this "videoconference fatigue" feeling. For instance, silence that naturally paces face-to-face conversations, is sometimes perceived during online communication as a technical degradation or interruption in the connection, which makes people feel anxious about the technology (Jiang 2020), and even judge negatively their interlocutor (Schoenenberg, Raake, and Koeppe 2014). Other conversational cues typically used in face-to-face interactions such as eye contact and turn taking are also degraded during video calls, rendering the conversation more difficult to follow and participate in. Moreover, the stress of being on camera while people stare at you could make the speaker to look at their own camera to be aware of their own behaviour, which again adds another focus of attention.

With all these challenges present, it is unknown whether having live transcriptions during video calls would be beneficial or detrimental for CI listeners. To the best of our knowledge, no research has investigated the performance of CI users during video calls when certain videoconference features are available (e.g., live transcriptions). To that end, we designed an online test to investigate CI users' preferences and listening efficiency during video calls under three presentation modes: audio (audio only), video (audio-visual), and captions (video plus captions). The study aimed to examine first which presentation mode was preferred by CI users compared to NH controls when attending to examples of pre-recorded online conversations. Secondly, participants' listening efficiency was calculated using a LBA model (Brown and Healthcote, 2008) that performs a joint analysis of behavioural responses (i.e., accuracy and response time). In this way, we compared the listening efficiency of CI and NH listeners when performing a behavioural test of speech recognition under the same three presentation modes.

Ultimately, this study aimed to investigate the cognitive demands perceived by CI users during online communication and which video call features (different presentation modes) may help them to communicate more effectively, taking into account not only their speech recognition but also the cognitive effort involved in doing so. Results of the study could provide the evidence base needed to offer appropriate advice to people with HL, and to everyone, including healthcare professionals, needing to communicate effectively with them using virtual platforms.

5.3 METHODS

The study received ethical approval by the North West - Greater Manchester Central Research Ethics Committee (REC reference: 20/NW/0141).

5.3.1 Participants

Two groups of participants were recruited: a group of CI users and a group of age-matched NH controls. All participants met the general inclusion criteria set as adults volunteers (aged 18 or over), who spoke fluent English (able to read and understand speech), with normal or corrected-to-normal vision (e.g., glasses), with capacity to give informed consent, able and willing to complete the study, and with no cognitive impairments (e.g., dementia). Additionally, participants in the CI group were required to have at least one CI, while those in the NH group confirmed that their hearing was good and that they did not use any hearing devices. Participants were able to access the test online, after reading the participant information sheet, confirming that they met the inclusion criteria as defined above, and providing informed consent. The test lasted approximately 30 minutes to complete and volunteers received a £5 pound gift voucher after their participation.

5.3.2 Distribution

The online test was open for participation from the 8th of October 2021 until the 7th of January 2022. The first recruitment target was the CI user group who were primarily recruited through the NIHR Nottingham BRC participant database. An invitation email containing the link to the test was sent to all CI users who were already members of the database. The test was also distributed by national and regional hearing charities and organisations in the United Kingdom including the Royal National Institute for Deaf People (https://rnid.org.uk/), the National Cochlear Implant Users Association (https://www.nciua.org.uk/), the British Association of Teachers of the Deaf (https://batod.org.uk/) and Hearing Link (https://hearinglink.org/).

Once the CI group recruitment finished, the recruitment of age-matched NH participants started. They were also recruited through the NIHR Nottingham BRC participant database and other national charities and organisations such as Age UK (https://ageuk.org.uk), the University of the Third Age (https://u3a.org.uk/) and the Life Cycle UK (https://lifecycleuk.org.uk). The test was also advertised within the School of Medicine and Health Sciences of the University of Nottingham.

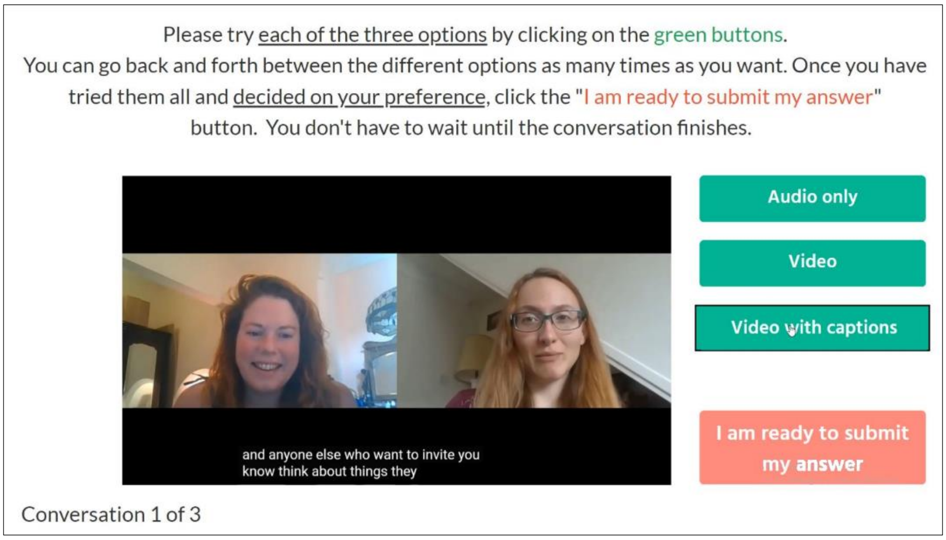
5.3.3 Test procedure

The test was implemented using Labvanced (www.labvanced.com), a web-based platform for online experiments (Finger et al. 2017), and consisted of three main sections: 1) demographic (age, gender, employment) and hearing information (hearing loss severity, hearing device use, and ways of communication in daily life); 2) video call preference task; and 3) main behavioural task.

Participants were required to answer all questions, although some items (conditional questions) in the demographic and hearing information section were only displayed when relevant (e.g., only participants who use a cochlear implant were asked about the years of experience using it). For a full reproduction of the demographic and hearing questions, see Appendix C: TEST ITEMS.

After the questionnaire, some set up instructions (Appendix C: TEST ITEMS) were given to participants to ensure an appropriate reproduction of the test on their computers or tablets. Participants in the CI group were instructed to use their hearing devices (CIs and/or HAs) as they normally do. All participants were asked to adjust the volume of their computers or tablets at a comfortable level (not too quiet, not too loud). They were also asked to indicate which sound reproduction settings (loudspeakers from computer or tablet, headphones, stream sounds to their hearing devices (only for CI users) or other), and button selection settings (touchscreen device or computer with mouse) they intended to use during the test. They were instructed to use the settings that they would normally use during a video call ("*Which sound reproduction setting are you using during the experiment? Please use the set up that you would normally use during a video call or the one you feel more comfortable with*"). Participants were required to use the same approach (their setting choice) throughout the experiment. Although this presentation setup is inherently associated with variations in the perceived loudness of stimuli between individuals, it ensures that stimuli are displayed in an ecologically valid way, making the results applicable to participants' real-world listening experiences.

Participants then accessed the second section (the preference task) that was designed to collect their preferences and opinions about different video call presentation modes. To do so, three examples of pre-recorded online conversations were presented to participants in random order, each of which was displayed in three different modes: audio only, video (audio-visual presentation), and video plus captions. Figure 5.1 shows one of the conversations displayed under the "video with captions" presentation mode. Participants were required to try each mode, switching at will between them, before deciding on their preference ("*which presentation mode would you prefer to use if they were real video calls that you were involved in?*"). There were no time restrictions, and therefore participants could spend as much time as they wanted viewing each mode. The conversation recordings were repeated in a loop until a decision was made (6 minutes length). Once participants tried each mode and were ready to submit their answer, three free-text questions were asked: i)"*If this was a real video call you were involved in, which mode would* *you prefer to use?*"), ii) "*Please tell us why it was your preferred mode*", and iii) "*Were there any downsides to your preferred mode?*"



Please try each of the three options by clicking on the green buttons.
You can go back and forth between the different options as many times as you want. Once you have tried them all and decided on your preference, click the "I am ready to submit my answer" button. You don't have to wait until the conversation finishes.

Conversation 1 of 3

Figure 5.1. *Screenshot of the test's second section (preference task) while conversation B was displayed under the "video with captions" presentation mode.*

Subsequently, participants performed a behavioural task of speech recognition, where both accuracy and response time were recorded under the same three presentation modes that act as experimental conditions. The task had 60 trials in total, 20 per condition, which were presented in random order.

Each trial comprised a sentence-length clip, on average 2.5 seconds (SD: 0.2s), immediately followed by the behavioural task (Figure 5.2). In total, 60 video clips, taken from an archive of BBC recordings, were displayed depending on the experimental condition either as an audio (visual information not provided, instead a loudspeaker icon was shown in the screen), or a video clip (with or without captions). Although the same video clips were presented to all participants, each participant had a different pseudorandom allocation of them to the three experimental conditions. Video clips were played only once at the beginning of the trial, i.e. there was no need for any button pressing to start the reproduction. After viewing them, participants attempted to select three keywords that were mentioned in the clip.

Each keyword was presented in turn alongside three alternative response options, one of which was semantically similar to the keyword, one of which was phonetically similar, and one of which was dissimilar. These four words (the keyword and the three alternative words) were displayed on the screen in random order in green buttons positioned along a vertical column. Once participants attempted to select the first keyword, another set of four words (the next keyword and its alternatives) appeared on the screen in parallel for participants to spot the next keyword. In total, three sets of four words (twelve words) were presented during the task. It is noteworthy that when a new set of words appeared, the previous sets were still visible on the screen, although in dimmed grey colour to indicate that the buttons were not active anymore. Participants' response time was recorded for each keyword selection from the moment in which the set of words appeared on the screen until the moment in which one word of the set was selected. There were no time restrictions, so participants could take as much time as they needed to select the words. Nonetheless, participants were specifically instructed to give equal weight to speed and accuracy when making their responses, and to give their best guess if unsure. As soon as the third keyword of a given trial was selected, the next trial started automatically. Participants were only allowed to take a break at mid-test, after having completed 30 trials. Prior to the task, participants completed a short practice session with six trials, two per condition, to become familiar with the task and stimuli.

It should be noted that participants were not informed about the accuracy of captions presented throughout the test, which was perfect on the behavioural task but moderate on the preference task.

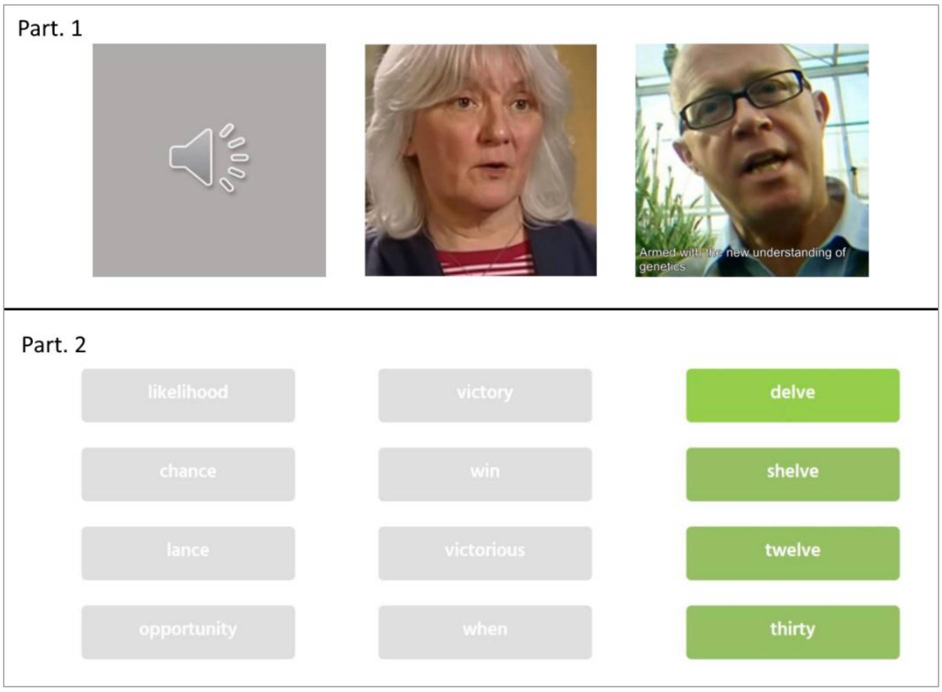


Figure 5.2. *Schematic representation of behavioural trials with video clips displayed first, followed by the behavioural task. Part1. Examples of sentence-length clips from the LRS2-BBC database displayed under the "audio", "video", and "video with captions" experimental conditions, respectively. Please note that all clips were displayed full screen for captions to be easily readable. Part2. Example of a behavioural task for the sentence: "when Lorna will have the chance to win up to twelve" (video clip ID: 5854842724793826670-00035). In the example, the first two keywords have already been selected, and the third keyword "twelve" is to be selected among its alternatives ("delve", "shelve", and "thirty").*

5.3.4 Materials and Stimuli

5.3.2.1 Video call preference task

The three conversation examples used in the preference task were recorded using Teams and Zoom platforms. Six adult volunteers, five native and one non-native English speakers, from the NIHR Nottingham BRC team provided consent for the audio-visual recordings to be used as stimuli in the present study. To simulate natural video calls, the conversations were unscripted. Nonetheless, each conversational couple were given a starting prompt topic that they could developed freely. Speakers were instructed to carry on with the conversation in the same way that they would normally do during a video call, even if any connection problems or technical difficulties arose. Details of the three conversations are given below:

- **Conversation A.** Speakers: female and male (non-native English speaker). Topic: "*Discuss which three personality traits/qualities you both like (admire) and dislike in a person and why*". Platform: conversation recorded using Microsoft Teams with a video frame rate of 8 frame per second (fps) and an audio sample rate of 16KHz. Quality: there were some connection problems and/or perceived difficulties including poor audio and video quality due to recording's low frame rate, bad lighting that partially obscured one of the speaker's face, and the presence of some speech overlapping during the conversation.
- **Conversation B.** Speakers: two females. Topic: "*Discuss and plan a dinner party that you would have together using only food and drink you both dislike*". Platform: conversation recorded using Zoom, with a video frame rate of 25 fps and an audio sample rate of 32KHz. Quality: one of the speakers' camera was slightly blurred.
- **Conversation C.** Speakers: two females. Topic: "*You both have won an all-inclusive holiday together (all expenses paid). Discuss and agree on where you want to travel to, the type of accommodation, for how long you will be there, and the activities you will book*". Platform: conversation recorded using Zoom, with a video frame rate of 25 fps and an audio sample rate of 32KHz. Quality: no connection problems or perceived difficulties were observed.

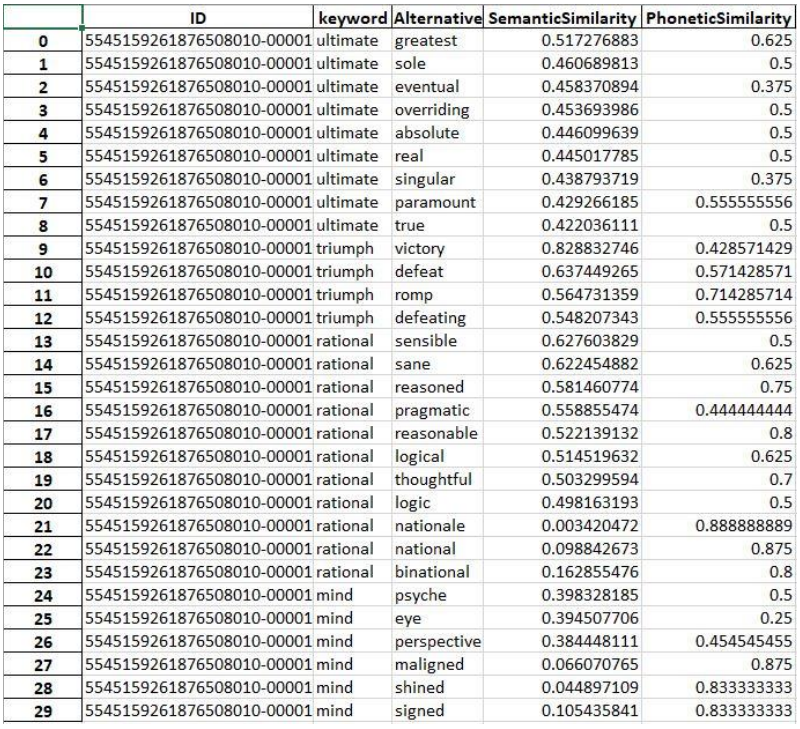
These conversations were selected as stimuli because they illustrate a range of different video call qualities that could naturally occur in real life. It is worth noting that the quality of the video call recordings was not intentionally manipulated. Captions for the three conversations were generated using Youtube automatic captioning, and were included in the "video with captions" presentation mode.

5.3.2.2 Main behavioural task

Sixty-six video clips from the Oxford-BBC Lip Reading Sentences 2 (BBC-LRS2) Dataset were used as stimuli in the main behavioural task (60 in the main task and six in the practice session). This database has previously been used in research for audio-visual speech recognition (Afouras et al. 2019; Chung et al. 2017; Son and Zisserman 2017). It consists of thousands of English spoken sentences, covering a wide range of topics, from the BBC television. A selection criterion was set to identify those videos that were, to a certain degree, like real video call conversations. For instance, the speaker should mainly look towards the camera, showing moderate head movements. Therefore, the selection criterion was primarily based on the speaker's gaze, head position, and overall movements. To apply these criteria, facial recognition analysis was carried out over the video clips of the main BBC-LRS2 dataset (48,165 video clips in total) using OpenFaceR toolkit as described by Cannata & Redfern, (2020). Parameters that control for the head position and gaze in both vertical (gaze_angle y=[-0.1,0.2]) and horizontal (gaze_angle x=[-0.25,0.25]) directions were limited around zero, which is the centred position. Similarly, parameters that control for head rotation movements in the three dimensions over the duration of the clip (sd(Rx,Ry,Rz)=[0,0.07]) were also constrained to a small range around zero, which indicates no movements at all. Video clips were further filtered based on other parameters such as video duration (between 2 and 3 seconds), video quality (bytes>=190.000), and word count (videos containing sentences with at least six words). In total 355 video clips met these criteria and were pre-selected to be used in the behavioural task.

The behavioural task was based on the sentences mentioned in the clips, whose transcriptions were available per video in text documents as part of the BBC-LRS2 Dataset. The identification of keywords within sentences was done in Phyton using spaCy (https://spacy.io/), an open-source natural language processing (NLP) package that classifies words according to their syntax (noun, adjective, verb, adverb, etc.). The candidate keywords considered were mainly nouns, verbs, adjectives, adverbs, and numbers (in words). The generation of alternative words per key word identified was based on the semantic and phonetic similarity. A list of ten semantically similar words were generated for each keyword using Gensim's Word2Vec model. The list of words generated was sorted in descending order based on their semantic similarity score. For instance, 'respectable' was the most similar word that the model found for the keyword 'decent' with a similarity score of 0.668, while 'modest' was in the list's thirteenth position, with a similarity score of 0.466. Phonetic similarity scores were also calculated for all the generated words (with respect to the keyword) using the "Metaphone" algorithm of the Pyphonetics Phyton library.

Another set of phonetically similar alternative words was generated for each keyword using the function 'get_rhymes' from the Phyton program "English to IPA" (https://pypi.org/project/eng-to-ipa/) that converts English text into the International Phonetic Alphabet (IPA). The phonetic and semantic similarity scores of these words with respect to the keyword were calculated using the Metaphone and Word2Vec algorithms respectively. Several checks ensured that all the generated alternative words existed in the English dictionary (using the "Brown Corpus" and the "word2vec-google-news-300" model), and that they were not the same as the original keyword (with different capitalisation, singular or plural form, containing the keyword, or same verb in a different form). If less than three alternative words passed the checks for a given keyword, then another set of ten alternative words was generated in an iterative process that repeated up to five times. A keyword was excluded if a minimum of three alternative words were not found after the checks. At the end of the process, a spreadsheet was exported for each pre-selected video, containing the list of candidate keywords with all its alternative words, and their semantic and phonetic similarity scores (Figure 5.3).



	ID	keyword	Alternative	SemanticSimilarity	PhoneticSimilarity
0	5545159261876508010-00001	ultimate	greatest	0.517276883	0.625
1	5545159261876508010-00001	ultimate	sole	0.460689813	0.5
2	5545159261876508010-00001	ultimate	eventual	0.458370894	0.375
3	5545159261876508010-00001	ultimate	overriding	0.453693986	0.5
4	5545159261876508010-00001	ultimate	absolute	0.446099639	0.5
5	5545159261876508010-00001	ultimate	real	0.445017785	0.5
6	5545159261876508010-00001	ultimate	singular	0.438793719	0.375
7	5545159261876508010-00001	ultimate	paramount	0.429266185	0.555555556
8	5545159261876508010-00001	ultimate	true	0.422036111	0.5
9	5545159261876508010-00001	triumph	victory	0.828832746	0.428571429
10	5545159261876508010-00001	triumph	defeat	0.637449265	0.571428571
11	5545159261876508010-00001	triumph	romp	0.564731359	0.714285714
12	5545159261876508010-00001	triumph	defeating	0.548207343	0.555555556
13	5545159261876508010-00001	rational	sensible	0.627603829	0.5
14	5545159261876508010-00001	rational	sane	0.622454882	0.625
15	5545159261876508010-00001	rational	reasoned	0.581460774	0.75
16	5545159261876508010-00001	rational	pragmatic	0.558855474	0.444444444
17	5545159261876508010-00001	rational	reasonable	0.522139132	0.8
18	5545159261876508010-00001	rational	logical	0.514519632	0.625
19	5545159261876508010-00001	rational	thoughtful	0.503299594	0.7
20	5545159261876508010-00001	rational	logic	0.498163193	0.5
21	5545159261876508010-00001	rational	nationale	0.003420472	0.888888889
22	5545159261876508010-00001	rational	national	0.098842673	0.875
23	5545159261876508010-00001	rational	binational	0.162855476	0.8
24	5545159261876508010-00001	mind	psyche	0.398328185	0.5
25	5545159261876508010-00001	mind	eye	0.394507706	0.25
26	5545159261876508010-00001	mind	perspective	0.384448111	0.454545455
27	5545159261876508010-00001	mind	maligned	0.066070765	0.875
28	5545159261876508010-00001	mind	shined	0.044897109	0.833333333
29	5545159261876508010-00001	mind	signed	0.105435841	0.833333333

***Figure 5.3**. List of candidate keywords and generated alternative words for the video clip ID: 5545159261876508010-00001, whose sentence is "The ultimate triumph of the rational mind".*

A final analysis was done in R (Version 4.1.2; R Core Team, 2021) to select only the 12 words needed in the behavioural task per video: three keywords and three alternative words per keyword. The selection of the three alternative words was done based on their semantic and phonetic similarity scores. If alternative words were to be represented as points whose coordinates (x,y) correspond to their semantic and phonetic similarity scores respectively (Figure 5.4), a perfect semantically similar word would have coordinates *(1,0)*. The perfect match in terms of phonetic similarity would be a word whose coordinates were *(0,1)*, and finally, a neither semantic or phonetic similar word to the keyword would be represented as *(0,0)*. These coordinates were used as reference points for the three 'ideal' alternative matches (reference point for semantic*: RS(1,0)*, phonetic: *RP(0,1)*, and neither: *RN(0,0)*). The three alternative words finally selected (the alternative semantic, phonetic and neither) for a given keyword were those words whose coordinates were at the shortest distance to the three corresponding reference points.

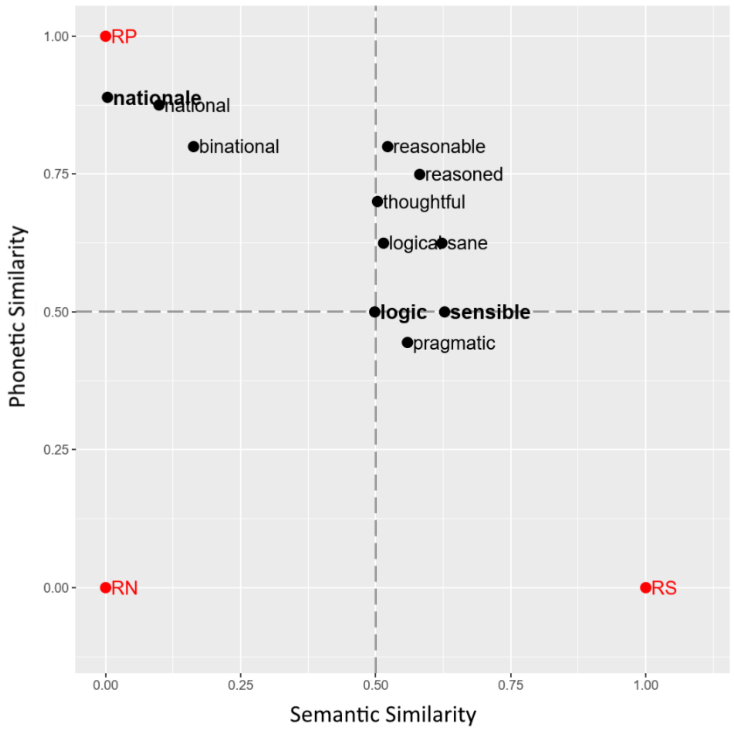


Figure 5.4. *Visual representation of all alternative words (by their similarity scores) generated for the keyword “rational” (video clip ID: 5545159261876508010-00001). The reference points for phonetic (RP), semantic (RS) and neither (RN) similarity are displayed in red. The three alternative words selected (closer to the three reference points) are in bold.*

Some checks were carried out to ensure that each of the three alternative words selected were valid and different from each other. If a given word did not pass the checks, the next closest word to the reference point was then selected in an iterative process. Keywords were excluded when less than three alternative words survived the checks. Likewise, video clips with less than three candidate keywords were excluded from the task. The keyword selection for those videos with more than three candidate keywords was done based on their syntax. Nouns and adjectives were prioritised since their alternative generated words were usually more accurate. A database with 265 video clips was created, containing the 12 words needed per video clip to be used in the behavioural task.

Captioned versions of all 265 videos were created to be used in the “video with captions” experimental condition. The scrolling text captions were generated in a custom script in Matlab, using the clips’ sentence transcriptions, and burned into the media file using FFmpeg (Version 4.4; FFmpeg developers, 2000-2021; http://ffmpeg.org/).

The final 66 video clips used in the main task (20 video clips per 3 experimental conditions and 6 additional clips for the practice session) were randomly chosen from the 265 pre-selected videos.

5.3.5 Analysis

Participants’ responses were exported as CSV (comma separated value) files from the Labvanced platform. The response data were anonymised, and participants were identified by a unique code. Separate databases were created combining all participants’ responses by section (demographic and hearing information, video call preferences, and main behavioural task) using RStudio software (Version 4.1.2; R Core Team, 2021).

Group level differences regarding video call preferences were analysed with Bayesian statistics using the brms R package (Bürkner 2017). Poisson regression models were fitted to examine the interaction of presentation mode by group (formula: Counts ~ Mode * Group) and the two-way interaction presentation mode by group and conversation type (formula: Counts ~ Mode * Group * Conversation). Default non-informative prior distributions were used, assuming that all values were equally likely a-priori. Posterior distributions were estimated using the MCMC(Van Ravenzwaaij et al. 2018) algorithms, whose convergence was measured by the potential scale reduction factor R-hat(R) (Brooks and Gelman 1998) over four separate chains, each with 2000 warmup iterations followed by another 2000 post-warmup iterations. Posterior predictive checks were performed to ensure that the models’ predictions adequately fit the data. The model conditional effects, predicted means and 95% credible intervals were reported per group and condition.

A hierarchical LBA model was performed to analyse participants’ behavioural responses following Gunawan et al. (2020). The LBA model was implemented in Stan as described in Annis et al.’s article (Annis et al. 2017), using a non-centred parameterisation to efficiently explore the posterior parameters’ distributions.

Seventeen free parameters were considered in the model: four evidence accumulators that represented each response option (correct, incorrect semantic, incorrect phonetic and incorrect neither) per each experimental condition (audio, video, and captions); two additional accumulators (vcBoots.Kw2 and vcBoots.Kw3) that were free to vary across conditions and explored the speed of evidence processing towards a correct response if the previous keywords were correctly answered; and finally, parameters A, b and t0 were able to vary between groups but were fixed across conditions. The model scaling constraint was set so that the between-trial variability in the drift rates was fixed to one (Sv=1). Uninformative priors (normal and Student-t distributions) were used for the model parameters. Posterior distributions were estimated using MCMC algorithm, over four separate chains, each with 1000 warmup iterations followed by another 2000 sample iterations (8000 draws in total). Posterior predictive checks were used to assess the agreement between model predictions and observed data.

Effects were assessed using 95% HDI (HDInterval package) that acts as the 95% credible intervals (CrI) of the mean posterior parameters. We considered the effect of group to be reliable when the 95% CrI of the between-group difference in posterior parameters (drift rates across conditions) did not contain zero.

Open (free-text) questions were analysed following Elo and Kyngäs’s guidelines for inductive content analysis (Elo and Kyngäs 2008) as described in Section 4.3.4. The resulting themes and categories identified were reported in tables per each conversation and group (Appendix D: THEMES AND CATEGORIES IDENTIFIED IN THE ONLINE TEST’S TEXT RESPONSES). Please note that participants’ text responses were analysed for informational purposes and that the study conclusions were based solely on quantitative results.

5.4 RESULTS

5.4.1 Participant demographics and hearing profile

A hundred participants completed the test, 50 per recruitment group. As shown in Figure 5.5, participants’ age in the CI group ranged from 20 to 86 years (M=64.2, SD=15.1), with 74% of them being female. The NH group showed a similar age range that spanned from 22 to 86 years (M=59.3, SD=12.2), with a more balanced gender distribution (54% of females participants). While most participants in the CI group were retired (62%), only 30% of them in the NH group were retirees. Most NH participants were currently working in part-time (32%) or full-time (30%) jobs.

All participants in the NH group reported having good hearing and confirmed that they did not use any hearing assistive technology. Nonetheless, 30% of them, those most senior (over 50 years old) reported having some degree of age-related HL.

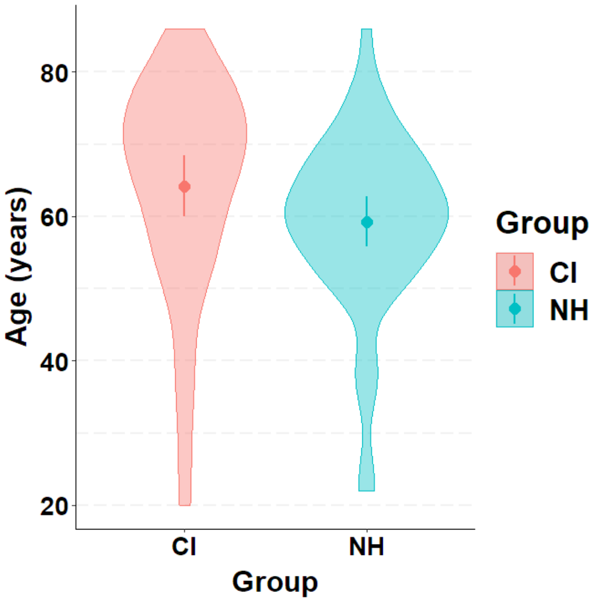


Figure 5.5. *Participants' age distribution by group. The groups' age mean is displayed by solid points with the lower and upper range representing the 95% confidence interval.*

Participants in the CI group had on average 10 years of experience using their implants (range from 2 to 26 years). The onset of HL in the implanted ear(s) occurred early in life (up to their teens) for 46% of participants, whereas 52% of them lost their hearing at middle adulthood (30 to 60 years old). Attending to the hearing device configuration, 58% of participants were unilateral CI users (one CI), 34% were bimodal users (one CI and a contralateral HA), and 8% were bilateral CI recipients (two CIs). Figure 5.6 shows CI users' hearing device configuration by age group. All participants with one CI (92%) reported having severe (6%) or profound HL (86%) in the non-implanted ear. Nearly all bimodal listeners (30%) reported using their HA regularly on a daily basis and wore it during the experiment (28%).

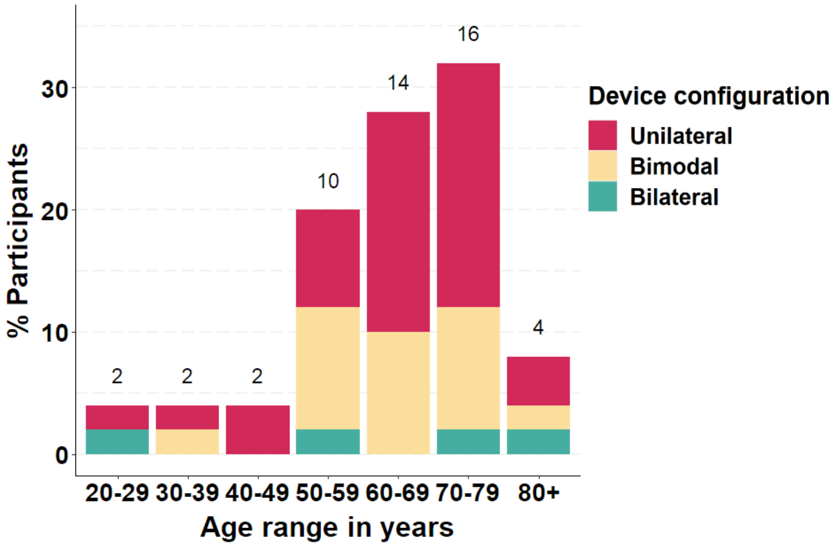


Figure 5.6. *CI users’ hearing device configuration by age group. Number of participants in each age group is shown at the top of each bar.*

Despite significant differences in participants’ hearing profile between both groups, common features can be seen in Figure 5.7 regarding their ways of communication in everyday life. Nearly all participants (88%) in both groups reported relying almost always on auditory speech for communication. Most participants in both groups reported making regular use of facial expressions to support communication, although CI users to a greater extent (56% of NH participants and 78% of CI users). CI participants also made regular use of other visual cues such as lipreading (76%) or text transcriptions (52%) to further support communication. Conversely, very few participants in the NH group reported using them (6% and 12% for lipreading and text transcriptions, respectively). Hardly any participants in either group reported using sign language to communicate with others (4% and 10% of participants used sign language “sometimes” in the NH and CI group, respectively).

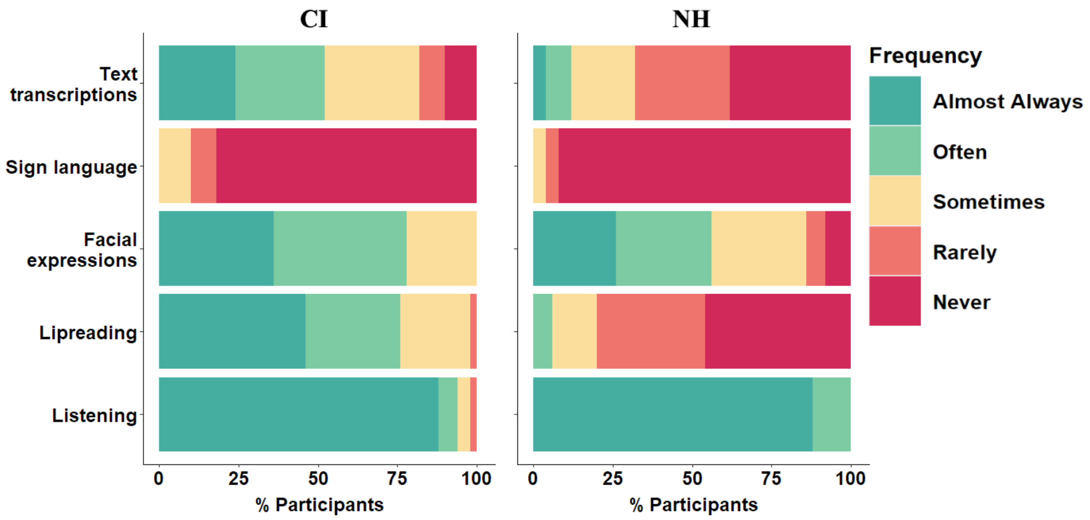


Figure 5.7. *Percentage of participants who rely on different ways of communication (listening, lip reading, facial expressions, sign language and text transcriptions) in everyday life by group. Q8: "In everyday life, to what extent do you rely on these ways of communication?"*

5.4.2 Participants' computer settings

Most participants in both groups completed the test using a computer with mouse (90% and 74% in the NH and CI group, respectively), as opposed to those using a touchscreen device (e.g., tablets).

When participants were asked to use the sound reproduction setting that they would normally use during a video call, most participants chose to use loudspeakers (72% and 68% in the NH and CI group, respectively). Others preferred the use of headphones to complete the test instead (28% and 12% in the NH and CI group, respectively), whereas a small portion of CI users (20%) streamed sound to their hearing devices.

5.4.3 Video call preferences

As shown in Figure 5.8, most participants in the CI group chose the Caption presentation mode for all three conversations (92%, 69%, and 63% for conversations A, B, and C respectively), whereas NH participants overall preferred the Video mode (64%, 80%, and 72% for conversations A, B, and C, respectively). Such differences were also evident as reported by the Poisson regression model results in the interaction presentation mode by group. No overlap in the model conditional effects and 95% credible intervals (CrI) indicated a large effect of group in participants' preferred presentation mode (Figure 5.9).

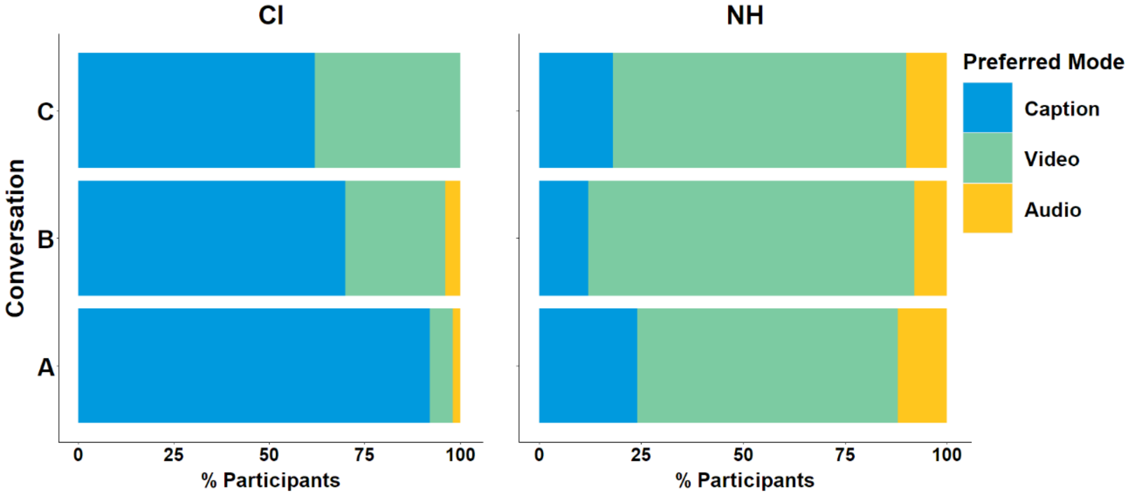


Figure 5.8. *Participants' preferred video call presentation mode per each conversation (A, B, C) by group.*

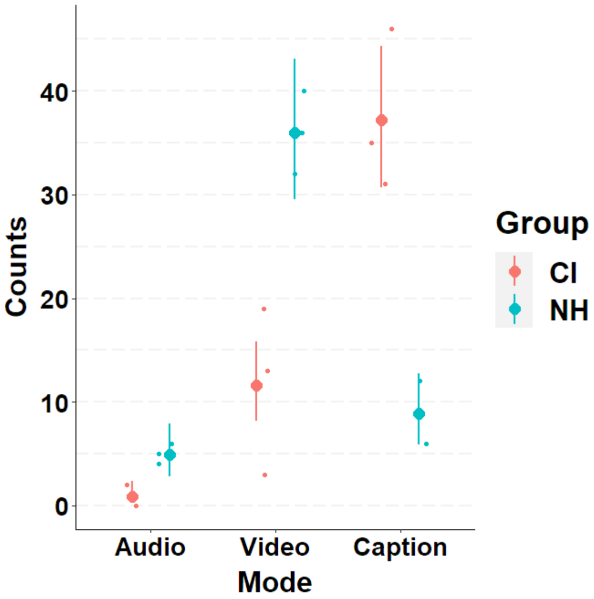


Figure 5.9. *Conditional effects plot of the Poisson regression model (Counts ~ Mode * Group) over participants' preferences raw data. The error bars display 95% credible intervals; the bold dots represent posterior medians, and the small dots represent the raw data (counts per each preferred mode per conversation by group).*

The analysis of open free-text questions revealed the reasons why participants selected their preferred presentation mode and any downsides. All resulting themes and categories identified are described in Appendix D: THEMES AND CATEGORIES IDENTIFIED IN THE ONLINE TEST'S TEXT RESPONSES.

According to CI users' comments, captions were overall preferred because they allowed them to understand all conversation, providing confirmation of what they heard and a backup in case of any missing word. They were especially useful when lipreading was difficult or not possible due to unclear speech, speakers' accent, fast speech pace, or even unfamiliar or technical topics. Captions were also considered helpful to overcome technical difficulties such as poor audio or video quality, and mismatch between the audio-visual content. Although most CI users were satisfied with the caption presentation mode, some downsides were also identified. The main problem mentioned was related with the accuracy and synchronization of captions. Participants reported that unsynchronised captions were distracting and could cause cognitive overload. The way in which captions were displayed on the screen was another issue reported. Participants in the CI group pointed out the importance of having them in big readable letters, with better punctuation, and different colours that facilitate the identification of different speakers. Likewise, the position of the captions at the bottom of the screen was considered a problem by many participants who reported being unable to read the captions and look at the speakers' faces at the same time. It prevented them from accessing visual information such as lip patterns and facial expressions. Those participants who tried to keep switching their gaze between the text and the speakers' faces during the conversation, ended up experiencing some degree of cognitive overload. Having the captions at a mid-screen position or even closer to the speakers' faces, in a speech bubble presentation, were solutions proposed to overcome this difficulty.

Conversely, participants in the NH group generally found captions distracting and not needed. Instead, most of them selected the video (audio-visual) presentation because they found it more natural and real, like a face-to-to-face conversation in which they could engage easily. NH participants found facial expressions useful to provide additional information that enhanced speech comprehension, such as identifying who speaks, the speakers' attitudes and reactions, and the overall tone of the conversation. A few participants reported to partially lipread even when the speech was clear. Most NH participants who chose video as their preferred mode did not report any downsides. Nonetheless, a few participants found that visual information sometimes could be distracting from the actual conversation (e.g., unfamiliar backgrounds). Likewise, a bad connection quality could also disrupt the conversation due to the presence of background noise or a lag in the audio-visual information. Although captions were not needed, some participants recognised their value in case of missing words, especially in the presence of accents, unclear or fast speech.

Although participants' preferences overall seemed to be independent of the conversations' quality, their comments revealed that they did notice differences in the quality between the three conversations that could have influenced to some extent the choice of at least some of them. For instance, the percentage of participants who chose the caption mode in conversation A was greater than in the other two conversations. This is true not only for the CI group but also for NH participants. Considering that conversation A was recorded under the worst conditions, it is not surprising that a greater number of participants opted to use captions to overcome technical difficulties such as the presence of background noise, poor video quality, speaker's accent and speech overlapping. Likewise, since conversations B and C were of better quality, more participants selected the Video mode for these conversations compared to conversation A. These patterns can be seen in the Poisson regression model results that examined the interaction of presentation mode by group and conversation type (Figure 5.10). Nonetheless, the effect of conversation type is small (not significant) relative to the uncertainty displayed by the 95% CrIs.

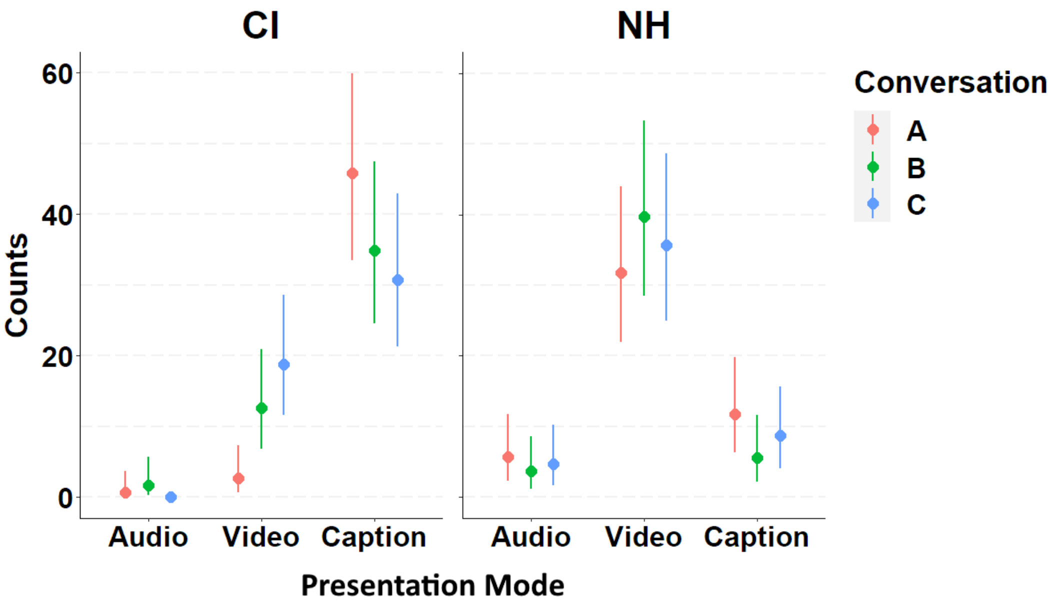


Figure 5.10. *Conditional effects plot of the two-way interaction Poisson regression model (Counts ~ Mode * Group * Conversation). The error bars display 95% credible intervals (CrI) and the bold dots represent posterior medians.*

5.4.4 Behavioural results

A total of 18,000 observations (60 trials x 3 keywords selection x 100 participants) were submitted to the LBA model. None of the observations was excluded due to unusual timings (extremely slow or quick responses). Satisfactory convergence was found for all estimated parameters as revealed by the full trace plots and Rhats' range [0.99-1.01]. Posterior predictive checks are shown in Figure 5.11 for each group and condition, for both correct and incorrect responses. Incorrect response types were summed up and plotted as negative. As can be seen, the model posterior predictive RT distributions followed closely the observed data. The difference between observed and predicted data in the NH group for all conditions was less than or equal to 1%, whereas in the CI group such difference increased slightly up to 1.8%.

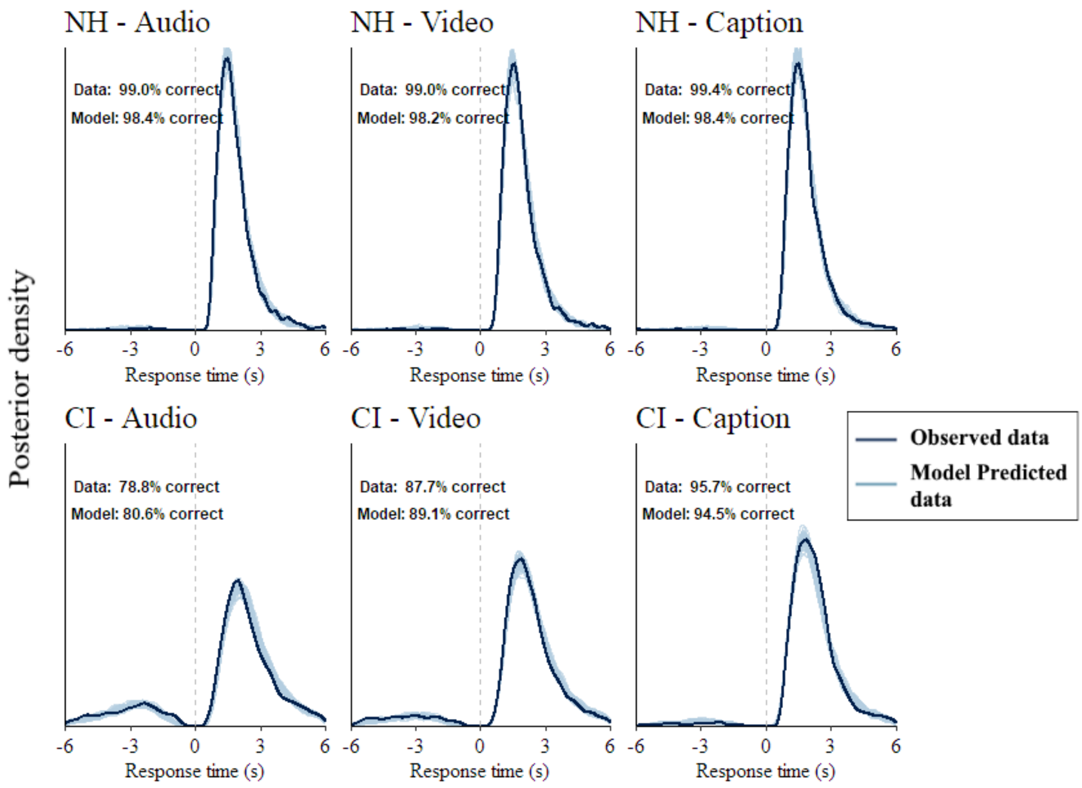


Figure 5.11. *Posterior predictive checks per group (NH and CI group shown at the top and bottom lines, respectively), and condition (audio, video, and caption mode shown from left to right columns). Participants' response time (RT) for incorrect responses are plotted as negative. Solid dark lines represent the observed data and light blue lines represent the model predicted data (8000 draws).*

Significant group differences were found in the posterior distributions of the LBA drift rates parameters (Figure 5.12). Particularly, the greatest contrast occurred in the drift rates of correct responses across experimental conditions (vCorrect.Audio, vCorrect.Video, and vCorrect.Caption). Overall NH participants showed considerably higher listening efficiency (i.e., faster accumulation of evidence towards correct answers) than CI users in all presentation modes. Such an effect of group was confirmed by the 95% CrI of the between-group difference in drift rates, which did not contain zero (Table 5.1). Interestingly, NH controls were equally efficient when performing the listening task regardless of the experimental condition (see correct responses' drift rates in Figure 5.12 and Table 5.1). Conversely, CI users' listening efficiency improved significantly across conditions, with credible intervals that hardly overlapped (Figure 5.12 and Table 5.1). These results suggest that unlike NH participants, CI users found accumulated benefits from the addition of visual cues and captions, respectively.

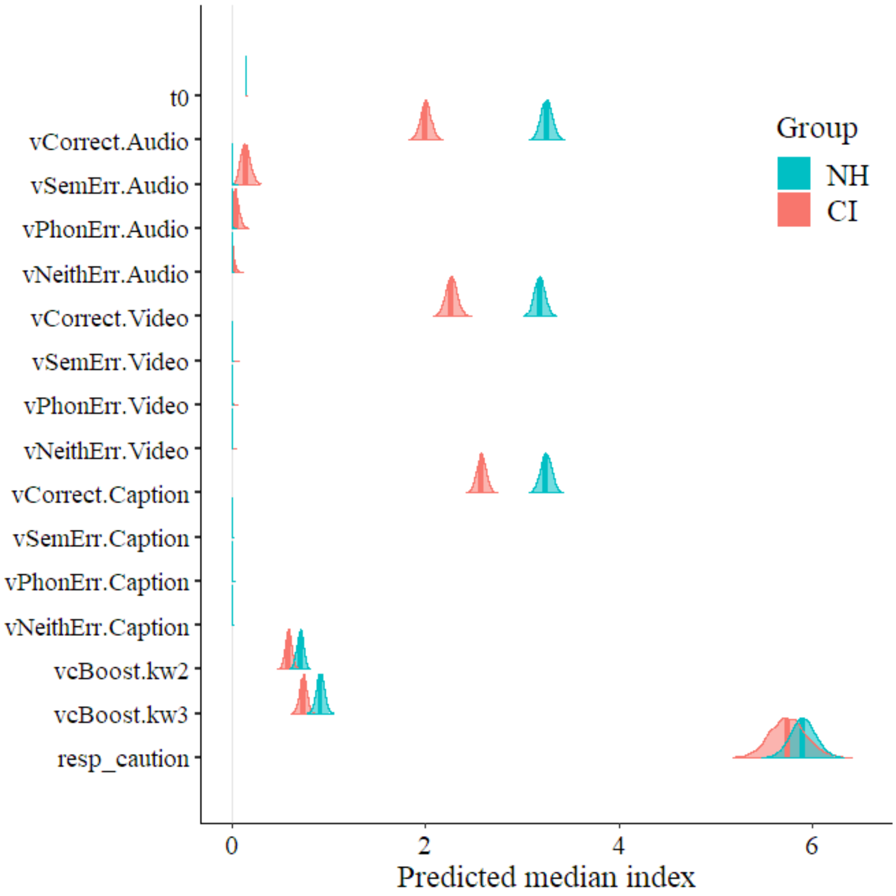


Figure 5.12. *Posterior group comparison in LBA model's parameters. Parameters are: t0, response caution (resp_caution), and differential drift rates (v) per response option (Correct, SemErr, PhonErr, NeithErr), and condition (Audio, Video, Caption), and two boost accumulators (vcBoost.) for Keyword 2 and 3 (kw2, kw3). Solid lines in posterior distributions represent the predicted median index for each parameter.*

The low frequency of incorrect responses (0.8% and 12.6% for NH and CI participants, respectively) did not provide enough data to perform an in-depth analysis of error types. NH participants could correctly recognise almost 100% of the keywords during the task in all experimental conditions, resulting in drift rates close to zero for all error type accumulators in all experimental conditions. Although CI users were also able to correctly recognise most keywords, especially in the Video and Caption conditions (showing drift rates close to zero too), they did make some errors in the Audio condition. Indeed, the lack of additional cues to support speech comprehension was evident in the audio only presentation mode where CI participants made most mistakes. This translated into CI users' drift rates towards incorrect responses greater than zero in the audio condition. Despite the low error rate, there was a tendency towards the semantic type of errors in the CI group. In the audio condition, CI users were more likely to select word by mistake that was semantically similar to the keyword mentioned in the clip (40.1% of semantic errors) than other phonetically (32.4% of phonetic errors), or non-similar words (27.5% of neither errors). Therefore, the semantic error accumulator was greater than the phonetic or neither ones in the audio condition (Figure 5.12 and Table 5.1). Nonetheless, this within-group effect was small and not significant as confirmed by overlapping CrIs. The between-group difference in mean drift rates for semantic errors however did show an effect of group in the audio condition (95% CrIs did not include zero), suggesting that CI participants made significantly more semantic errors than their NH peers.

The two additional accumulators (vcBoots.Kw2 and vcBoots.Kw3) included in the model confirmed that having understood precedent keywords facilitate the recognition of subsequent keywords. Participants in both groups showed faster accumulation of evidence towards a correct answer (increased efficiency) when the previous keyword in the sentence was correctly identified, and even faster when the two previous keywords were right. These results suggest that the context (understanding the precedent part of the sentence) enhances listening efficiency. However, this listening efficiency improvement was greater in NH participants than in CI users as revealed by the 95% CrI of the between-group difference drift rates (that did not contain zero).

No effect of group was found in other model parameters such as the non-decision time (t0) and response caution (calculated as K+A/2). This indicates that the amount of evidence required to trigger a response was not significantly different between groups, showing similar levels of caution. Likewise, the time taken for other non-decision processes (t0) such as perception (stimulus encoding) and response execution (motor response) was almost the same for participants in both groups, suggesting that the observed differences in response time between groups during the task were due to decision-making processes.

Table 5.1. *Means and Highest Density Intervals (HDI) of the LBA model's drift rates (v) posterior distributions by group. The effect of group can be seen in the between-group difference shown in the three right columns. The HDI is interpreted as the 95% credible intervals (CrI) of posterior distributions.*

	NH Group			**CI Group**			**Between-Group Difference *(NH-CI)***		
	Mean (v)	**95%CrI Lower**	**95%CrI Upper**	**Mean (v)**	**95%CrI Lower**	**95%CrI Upper**	**Mean (v)**	**95%CrI Lower**	**95%CrI Upper**
t0	0.1502	0.1500	0.1509	0.1502	0.1500	0.1513	-0.0001	-0.0015	0.0011
vCorrect.Audio	3.2524	3.1125	3.3904	2.0045	1.8769	2.1395	1.2479	1.0585	1.4391
vSemErr.Audio	0.0050	0.0000	0.0247	0.1461	0.0394	0.2582	-0.1410	-0.2596	-0.0346
vPhonErr.Audio	0.0063	0.0000	0.0298	0.0557	0.0001	0.1264	-0.0493	-0.1309	0.0095
vNeithErr.Audio	0.0034	0.0000	0.0157	0.0256	0.0000	0.0747	-0.0222	-0.0789	0.0062
vCorrect.Video	3.1820	3.0491	3.3153	2.2711	2.1282	2.4205	0.9109	0.7113	1.1083
vSemErr.Video	0.0018	0.0000	0.0090	0.0045	0.0000	0.0237	-0.0027	-0.0260	0.0127
vPhonErr.Video	0.0026	0.0000	0.0119	0.0064	0.0000	0.0264	-0.0039	-0.0289	0.0111
vNeithErr.Video	0.0017	0.0000	0.0086	0.0041	0.0000	0.0187	-0.0024	-0.0194	0.0099
vCorrect.Caption	3.2479	3.1131	3.3836	2.5813	2.4661	2.7026	0.6666	0.4825	0.8409
vSemErr.Caption	0.0010	0.0000	0.0051	0.0010	0.0000	0.0049	0.0000	-0.0060	0.0064
vPhonErr.Caption	0.0012	0.0000	0.0062	0.0010	0.0000	0.0054	0.0002	-0.0062	0.0075
vNeithErr.Caption	0.0009	0.0000	0.0042	0.0007	0.0000	0.0037	0.0001	-0.0046	0.0051
vcBoost.kw2	0.7076	0.6255	0.7938	0.5865	0.5056	0.6714	0.1211	0.0008	0.2358
vcBoost.kw3	0.9159	0.8135	1.0133	0.7382	0.6448	0.8263	0.1777	0.0384	0.3107
resp_caution	5.9056	5.5933	6.2254	5.7549	5.3132	6.2325	0.1507	-0.3677	0.7154

Table parameters are t0, drift rates (v), and response caution (resp_caution). Drift rates are described in the format "response option. Condition". Response options are: correct, semantic error (SemErr), phonetic error (PhonErr), and neither error (NeithErr). Conditions are: audio, video, and caption. Additionally, two boost accumulators (vcBoots) are defined for Keyword 2 (kw2), and Keyword 3 (kw3).

5.5 DISCUSSION

The study compared the subjective preferences and listening efficiency of a group of 50 CI users and a group of 50 NH controls when attending to pre-recorded online video call conversations and sentence-length video clips, respectively, under three presentation modes (or experimental conditions): audio, video (audio-visual), and captions (video plus captions).

Captions: a preference and a need for CI users

A significant effect of group was found in the participant preferences regarding video call presentation modes. As expected, CI users overall preferred the caption presentation mode while the choice of most of their NH peers was the audio-visual presentation (video mode). This was true for all three conversations regardless of the quality at which they were recorded. The availability of captions was essential for most CI users, even for those who having good speech understanding performance still needed them as backup information in case of missing words. Captions provided confirmation and reassurance of what they heard, which was particularly important under challenging conditions such as unclear speech, when the speaker has an unfamiliar accent, or even when technical difficulties arise (bandwidth problems, poor audio or video quality, etc.). Participants in both groups confirmed what previous research (Jensema et al. 2000) has already shown, that the presence of captions made them shift their attention form the visual content (the speaker´s faces) to the text content (live transcriptions). This shift in attention was somehow perceived as a tiring and distracting process that prevented lipreading and facial expression recognition. Although this disadvantage of captions did not impede CI users from obtaining some benefits from them, such distraction was not worth it for NH participants who were perfectly able to understand the conversation without the additional support of captions. Nonetheless, some NH participants recognised captions’ value under challenging listening situations (technical difficulties, etc.). Most participants in both groups agreed on the importance of captions’ accuracy and synchrony. Such precision was considered key for captions to provide benefits; otherwise, they could cause more confusion than support. However, live transcriptions seemed to be useful for CI users, even with moderate levels of accuracy (accuracy is typically around 60-70% for Youtube automatic captioning). These are reassuring results considering that most freely available captioning services have accuracy levels that can be even lower than those provided by Youtube. In this regard, more research is needed to assess how the perceived benefits of subtitles may vary under different levels of accuracy.

Interesting suggestions were also proposed by participants in the CI group to improve caption presentation. For instance, a closer position of the captions to the speakers´ faces would facilitate a quicker switch of gaze between media contents, potentially reducing the cognitive load associated with the attention shifting. Different colours of subtitles for different speakers could also enable a faster identification of who is speaking. According to CI users, even small changes in the captions' display mode such as using larger fonts and better punctuation would make a great impact in aiding speech understanding while reducing the cognitive cost.

CI users' listening efficiency enhanced by visual cues

Results of the study showed that CI users' subjective preferences mirrored their performance on the behavioural task. They not only perceived a subjective benefit of captions but they also performed better under the video with captions presentation mode. Indeed, the model results confirmed that CI users benefitted significantly (i.e., higher listening efficiency) from the addition of video content and benefitted further from the addition of captions. Such improvement in listening efficiency across conditions was significant within the CI group as revealed by not overlapping 95% credible intervals (CrI) of correct responses drift rates' posterior distributions. These results corroborate the assumption that the addition of visual cues, such as video and captions, enhances speech understanding for people with HL. In contrast, NH participants achieved similar levels of performance in all presentation modes as shown by almost identical posterior distribution of correct responses' drift rates across conditions. Interestingly, although NH controls reported captions being distracting, their performance in the caption condition did not reflect any difference compared to the other conditions. Nonetheless, we are unable to know which information they attended when completing the task under the caption condition. They could have ignored captions completely in case of any perceived distraction and focused only on the audio-visual content. It would be useful to learn whether captions can provide benefits to NH participants under less favourable listening condition, such as bandwidth-related limitations in speech quality, stuttering video streams, and potential lags between audio and video streams.

The difference in performance between participants in both groups was remarkable; a large effect of group was found in participants' listening efficiency (in correct responses) in all presentation modes. Overall, NH participants exhibited significantly better listening efficiency than CI users, as revealed by the 95% CrI of the between-group difference posterior distributions. This was true even under the captions condition where CI users accomplished better results. Yet their listening efficiency remained considerably below that achieved by NH controls. These results suggest that online communication may be more cognitively demanding for CI users than for their NH peers. This is not surprising considering that previous research has already shown that listening to speech is more cognitively demanding for CI users than for people with NH (Perreau et al. 2017), even when optimal intelligibility is achieved (Pals et al. 2020; Winn et al. 2015).

The importance of the context

The importance of the context was evident. Participants showed greater listening efficiency when they correctly identified the preceding keyword, and to a greater extent, when they got the two previous keywords of the sentence right. Such effect suggests that the context plays an important role in speech understanding (Dingemanse and Goedegebure 2019; Eisenberg et al. 2002; Sheldon, Pichora-Fuller, and Schneider 2008; Wilson and Dorman 2008). Previous research has already shown that having understood the first words of a sentence increases intelligibility scores for subsequent words. For instance, Winn and Teece (2020)'s study (Winn and Teece 2020) on 21 CI users showed that in high-context sentences, intelligibility scores for the final words reached 97.8% when preceding words were repeated correctly, whereas the accuracy of last words dropped to 56.5% when there was at least one preceding error. In line with this, our results showed that participants in both groups benefited from the context to increase their listening efficiency, although this effect was greater in the NH group. Although the benefits provided by semantic context are perceived by both groups (O’Neill et al. 2021), NH listeners continue to be more efficient in processing this information. This is not surprising given that they consistently showed higher listening efficiency in all presentation modes compared to CI users.

Likewise, despite the low rate of errors, there was a tendency in the CI group towards making semantic types of errors in the audio condition. When CI users were unsure about what word they heard in the audio presentation mode, they were more likely to select words that were semantically similar to the keyword, rather than words that were phonetically similar or not similar at all. Again, the context seems to play a role in the deduction of misheard words that may be more relevant than phonetic similarity. This idea has recently been proposed by Winn and Teece (2021) in a study that found the prevalence of errors driven by the semantic coherence was higher in high-context sentences. In their study, when participants were asked to repeat back what they heard, they sometimes replaced a misperceived word by another sensible word that fitted within the semantics of the sentence but that was not consistent with the phonetics of the misheard word. Moreover, the authors found using pupillometry techniques that linguistic coherence errors elicited smaller pupil dilations than actual correct responses when their semantic fitted well within the sentence. They concluded that LE may be a function of coherence reconstruction within listeners‘ perception rather than a function of the number of errors made. According to this, HI listeners would be more likely to make semantic errors given that the effort of coherence reconstruction is reduced because semantic errors still make sense within the sentence´s context. The way in which intelligibility was measured in this study differs considerably from Winn and Teece´s task in that participants didn´t have to repeat back what they heard but instead they attempted to select the keyword among four options. This task is easier since there was no need for participants to come up with a sensible word for the replacement of misheard words. Although this could have levelled up the chances of making any type of error, yet CI users still selected the semantically similar option more often (40.1%) than the phonetic (32.4%) or the non-similar one (27.5%). These results suggest that there is a search for semantic coherence that may be driving participants´ speech perception. However, given the low frequency of errors made during the task, an in-depth statistical analysis of error types was not possible, and thus, these findings are not conclusive. Future studies can replicate this experiment but manipulating the level of difficulty (e.g., SNRs) to investigate whether semantic errors are predominantly made by CI users.

Listening efficiency metric sensitivity

In this study, the listening efficiency metric was used to evaluate participants' performance when attending to video clips (simulating video calls), considering not only their speech intelligibility but also the cognitive effort involved. Results showed that this metric was able to capture significant differences between groups of participants and across conditions in the CI group. Indeed, remarkable differences in listening efficiency were found between CI and NH listeners in all experimental conditions. Moreover, despite CI users' intelligibility being at ceiling levels (low rate of errors), listening efficiency showed sensitivity to small changes in task demands between audio, video, and caption conditions. These results are consistent with CI users' subjective perspectives as they report online communication to be cognitively taxing and visual cues beneficial to aid speech performance.

Although no significant differences were found in the listening efficiency scores of NH participants across conditions, this seems to reflect similarities in performance rather than a lack of sensitivity. It is plausible that the three presentation modes are perceived equally effortless by NH listeners, given that none of them posed any listening challenge. Certainly, speech was clear in the three presentation modes, with no listening adversities that could have increased task demands. Nonetheless, further investigation is needed to confirm the sensitivity of this measure in the NH community under more challenging listening conditions.

Moreover, the listening efficiency metric showed sensitivity to the benefits gained from semantic context in both groups of participants. It is indeed reassuring that the efficiency metric was able to pick up an effect that has been already proved by previous research in both CI and NH listeners (O'Neill et al. 2021; Winn and Teece

2021). In addition, and despite the low error rates, this metric was able to indicate a tendency in the error typology made by CI users.

These results are promising and demonstrate that listening efficiency is capable of evaluating listeners' performance in realistic situations, even when intelligibility is optimal and changes in task demands are subtle. In addition, its application has been proved suitable even for behavioural tasks that are performed online.

5.5.1 Limitations

Participants volunteering to complete the online test were recruited online via email and social media. This could have introduced selection bias since those participants may be more familiar with online technology and thus with online communication. Despite both groups of participants being approximately age-matched, a greater number of CI users were retired compared to their NH peers (62% and 30% of retirees respectively). Such imbalance might have contributed in part to the observed difference in performance between both groups in the behavioural task, based on the assumption that workers may communicate online more frequently than retirees might.

Another limitation of the study is related to its online nature, which prevented us from having a greater control over experimental factors such as the reproduction settings in participants´ computers. To reduce this limitation, participants received detailed set up instructions and reported their selected settings. Although testing participants in a non-controlled environment is a limitation from the experimental point of view, on the other hand, it made the test more realistic or "ecological" because participants used the exact same devices and settings they would normally use during online conversations.

The stimuli used during the test added extra limitations to the study since were intended to represent online conversations but without allowing participants to interact. While the conversations used during the "preference task" were real pre-recorded online conversations, the sentence-length video clips used for the behavioural task were far from mimicking a real conversation since their goal was to deliver a listening test instead. Nonetheless, the BBC-LRS2 clips Dataset includes high context sentences that covered a wide range of topics, with different speakers, and from different locations, which somehow added more "realism" compared to other formal datasets with fixed-structure sentences commonly used for speech intelligibility tests (e.g., BKB sentences). Additionally, the captions used in the behavioural task did not contain any errors since they were generated from the sentences´ transcriptions. Such ideal conditions in captions´ accuracy is likely to have contributed to the increase in "listening efficiency" performance observed in the CI group (caption mode).

Finally, the low frequency of errors made by participants during the behavioural task did not provide enough data to perform a conclusive error type analysis. This prevented us from making inferences or drawing conclusions about the type of errors commonly made by CI users.

5.6 CONCLUSIONS

The presence of captions during video calls, together with audio-visual information, was the presentation mode overall preferred by CI users. They were considered a useful tool that allowed participants in the CI group to follow the conversation easily. Such impression was also confirmed by the behavioural results. A Hierarchical LBA model, a joint analysis of accuracy and response time, showed that CI users benefit significantly from the addition of visual cues (i.e., being able to see the talker's face) and benefit further from the addition of captions. Nonetheless, even with captions, CI users' listening efficiency remained significantly below that achieved by NH controls in all presentation modes. This suggests that online communication may be more cognitively demanding for CI users than for their NH peers, as revealed by slower and less accurate responses to what they have heard. NH controls, in contrast, achieved similar performance regardless of presentation mode. Given the benefits that captions provide to the CI community, and presumably more widely to the HI community, we considered that online communication platforms should make them available and adjustable according to individual need and preference. Improvements in captions' display mode are still needed to overcome downsides such as the attention shifting that occurs between video and text content, the difficulty in identifying different speakers, and problems with their accuracy and synchrony.

CHAPTER 6. PROJECT DISCUSSION

6.1 PROJECT SUMMARY AND AIMS

Over the past few decades, there has been an increased awareness that listening is more than our ability to hear, it involves auditory-cognitive interactions (Pichora-Fuller et al. 2016). Expressly, real-life listening is not only about hearing to an attended target correctly, but also about the amount of effort expended in doing so. This is particularly relevant in noisy environments, which increases the cognitive demands of the listening task (CHABA 1977). Although these challenges are commonly encountered in a noisy world, they may be exacerbated when the listener also suffers from HI. CI recipients, for instance, have severe HL and the impoverished auditory signal that they receive from their implants has been shown to tax cognitive resources during speech perception (Başkent 2012; Winn et al. 2015). Indeed, previous research has found that CI users report experiencing high levels of LE and fatigue in everyday life (Alhanbali et al. 2017). It is well known that this elevated cognitive load could affect negatively their life, particularly concerning their social participation (Kramer et al. 2006; Mick et al. 2014; Pronk et al. 2011; Shukla et al. 2020), and long-term cognitive health (Lin et al. 2013; Pichora-Fuller et al. 2015).

In clinical practice, the only way to evaluate these difficulties is by means of standardised listening tests that only assess the speech understanding performance of patients, without considering the cognitive dimension of listening. The most common are speech-in-noise tests (e.g., QuickSIN, BKBSIN,HINT) that measure the speech reception threshold (SRT; Plomp & Mimpen, 1979). This score usually indicates the level of background noise (SNR) that a patient can tolerate while being able to understand 50 percent of the words presented. Although these tests certainly provide useful information about the speech recognition performance of patients in background noise (Wilson, McArdle, and Smith 2007), they may not be entirely representative of real-world listening situations.

Unsurprisingly, in real life people with HL spend more time in less noisy environments (Wu et al. 2018), where they can understand most, if not all, of what they hear. It is in those situations when the assessment and understanding of LE become more relevant. Indeed, listening may be highly taxing for CI users even when their intelligibility remains high (Pals et al. 2020; Winn et al. 2015; Winn and Teece 2022). Hence, the importance of evaluating suprathreshold (i.e., above moderate intelligibility) LE in the CI community.

This issue formed the foundation of this thesis. The overall aim of the thesis was to investigate the LE experienced by CI recipients in realistic sound scenarios. We focussed on those listening situations that were especially relevant to social interactions. Specifically, having conversations in a busy atrium café and communicating through video call, situations that have become very frequent in everyday life. Likewise, the Covid-19 outbreak brought new communication scenarios that also deserved our attention given the severe difficulties reported by the HL community.

To obtain a comprehensive assessment of the mental exertion of CI users in the aforementioned situations, we used multimodal measures of LE, which included subjective, behavioural, and physiological assessments. Nonetheless, our focus was primarily on objective measures since they are not subject to individual bias and thus, are good candidate for being implemented into clinical practice. In this regard, we proposed two approaches to objectively quantify LE in the implanted population. Firstly, we explored the suitability of combining fNIRS-based brain imaging with simultaneous pupillometry as a research tool to investigate the neural correlates of effortful listening. Secondly, we proposed a joint analysis of behavioural responses (accuracy and response time) capable of providing a listening efficiency measure that reflects both intelligibility and LE. Regarding the latter, we further explored the sensitivity of listening efficiency to changes in cognitive demands, its relationship with other measures of LE, and its suitability for becoming a CI outcome measure.

A summary of the findings from each study is described in the following section. Then, an overall project discussion is presented. The chapter ends with an exploration of project limitations, impact, future directions, and conclusions.

6.2 SUMMARY OF FINDINGS

6.2.1 Main findings of Chapter 2

1) The drift rate parameter of a LBA model of choice decision-making can be used as a putative measure of listening efficiency. This outcome measure can reflect objectively both participants’ intelligibility and LE while performing behavioural listening tasks.
2) CI users were disproportionally affected by moderate ecologically relevant levels of background noise compared with NH controls, as revealed by much lower listening efficiency, even when their intelligibility was optimal at highly favourable SNRs. This confirms that LE can be experienced even when speech understanding performance is at ceiling levels.
3) CI users also reported considerably greater levels of perceived LE than their NH counterparts, both in daily life experiences and while performing the laboratory speech in noise task. Nonetheless, their ratings of task disengagement were consistently low across all experimental conditions.
4) Listening efficiency was sensitive to changes in task demands (within-subjects) and between groups of participants (between-subjects), even when intelligibility remained optimal.
5) Listening efficiency showed within the CI group (but not the NH group) moderate-to-strong correlations with cognitive scores and self-reported ratings of LE, both in the laboratory and in daily life. Thus, better cognition (i.e., working memory) and more positive listening experiences may be predictors of better listening efficiency in CI users.

6.2.2 Main findings of Chapter 3

1) The simultaneous acquisition of fNIRS brain imaging and pupillometry is a feasible approach to investigate LE in both NH and CI listeners.

2) CI users showed significantly greater levels of arousal compared with NH controls when listening to speech under ecological levels of background noise, as revealed by larger pupil dilations at baseline and shallower slopes of recovery to baseline. This is suggestive of high cognitive effort during noise exposure and increased need for recovery during active listening.
3) fNIRS and pupillometry revealed the same task-evoked cognitive load. LIFG activity, across conditions and groups, mirrored the PPD and MPD pupil results. However, these task-evoked physiological measures did not capture significant differences in LE between groups of participants.
4) Care must be taken when analysing task-evoked physiological measures in isolation as they may overlook any underlying cognitive processing (as revealed by pupil results at baseline) that can compromise task resource allocation.
5) Both physiological measures showed moderate associations with listening efficiency scores in the CI group. Therefore, pupil metrics (SL and BL) and brain activity in bilateral STG and RDLPFC may act as predictors of CI users' performance.

6.2.3 Main findings of Chapter 4

1) CI users considered that the introduction of facemask and social distancing during the COVID-19 pandemic significantly worsened their in-person listening experiences compared to pre-pandemic times. Moreover, the frequency with which they held telephone and video calls increased considerably during the pandemic.
2) CI users reported experiencing a diverse array of listening difficulties during the pandemic, including reduced speech intelligibility and increased LE. Ratings of LE were consistently high for both in-person and remote communication, regardless of the level of speech understanding achieved.
3) These difficulties lead to many CI users actively avoiding certain communication scenarios at least some of the time.

4) Visual cues were considered important to aid CI users' speech understanding. Solutions that offer improved access to visual cues were overall preferred, including transparent facemask or visors, video calls instead of telephone calls, and live speech-to-text subtitling.

6.2.4 Main findings of Chapter 5

1) Online communication using video calls was more cognitively demanding for CI than for NH listeners, as revealed by significantly lower listening efficiency in all presentations modes (audio, video, and captions).
2) Visual cues enhanced CI users' listening efficiency during online video calls. CI recipients benefited significantly from the addition of video (i.e., being able to see the talker's face) and benefited further from the addition of captions.
3) NH controls achieved similar performance (i.e., listening efficiency) regardless of presentation mode.
4) Performance on the behavioural task mirrored the subjective preferences. Overall, CI users preferred the caption presentation mode since it allowed participants to follow the conversation easily. Conversely, NH participants preferred the video mode as captions were perceived as distracting and not needed.
5) Participants in both groups benefited from the context to improve their listening efficiency. They were more efficient at identifying keywords when they understood correctly the first part of the sentence (previous words correct). Likewise, CI listeners seemed to rely on the context for the deduction of misheard words, given that semantic errors were made more often than other types of errors.

6.3 DISCUSSION

During the course of this project, we carried out three studies to assess the LE experienced by CI users in real-world scenarios. The first laboratory-based study used a variety of measures to estimate, both objectively and subjectively, the mental exertion of a group of CI users while they listened to speech under ecological relevant levels of “cafeteria” background noise. A survey was conducted to explore the listening difficulties perceived by CI recipients under communication scenarios commonly experienced during the Covid-19 pandemic. Finally, given the increased reliance on remote communication, we examined the cognitive demands that video calls posed to the CI community. Implications of the findings corresponding to each of these studies are explored in subsequent sections.

6.3.1 A neural marker for real-world effortful listening in CI users remains elusive.

Previous research has suggested that the recruitment of non-auditory prefrontal areas, particularly the LIFG, is associated with adaptive control mechanisms that optimise speech recognition in challenging listening conditions (Eckert et al. 2016; Wild et al. 2012; Zekveld et al. 2006). The engagement of this top-down mechanism is expected to be prevalent among CI users given the perceived listening difficulties imposed by their implants. Although such hypotheses have been confirmed by previous studies (Jiwani et al. 2016; Petersen et al. 2013; Sherafati et al. 2022; Strelnikov et al. 2015), very little is known about the frontal activity patterns of CI users in more realistic sound scenarios. Our laboratory experiment aimed to examine this by measuring the brain activity of CI recipients while they listened to speech in a real-world cafeteria background noise at ecologically relevant SNRs. However, contrary to our expectations, the fNIRS brain imaging results revealed similar cortical activity patterns in both groups of participants (CI and NH listeners). Several hypotheses were discussed in section 3.6 (CHAPTER 3) as potential explanations to justify the observed results. Among the most plausible explanations was the inability of finding common and representative brain activity patterns in the CI group due to the variability in their speech understanding performance. Although there is evidence that cortical activation may differ considerably between CI users as a function of their speech perception ability (Olds et al. 2016), we cannot confirm this interpretation given that fNIRS measurements were analysed at a group level. Another probable explanation is that complex realistic environments may pose higher demands that could have approached the limits of CI listeners' cognitive capacity. This could explain the drop observed in the LIFG's task-evoked responses during the hard condition, which also coincided with the task-evoked pupillary results (PPD and MPD). This rationale, although highly likely, cannot be confirmed with the current data. The experimental design did not considered the acquisition of fNIRS measures during the background noise exposure (baseline), which prevented us from obtaining brain-imaging evidence of any underlying cognitive load experienced by CI users. Further research needs to examine more closely these hypotheses to elucidate whether a neural marker of LE can be found in the implanted population under more realistic listening environments.

6.3.2 The suitability of combining fNIRS brain imaging and pupillometry techniques to investigate neural correlates of LE.

In this project, the combination of fNIRS brain imaging and pupillometry was proposed as a more comprehensive approach to investigate the neural correlates of effortful listening in CI recipients. Such a proposal was promising given the close correspondence of both techniques in targeting the cognitive resource allocation needed to optimise speech understanding performance (Aston-Jones and Cohen 2005; Zekveld et al. 2014). From a practical perspective, the compatibility of both physiological measures with CI devices and the possibility of using them simultaneously made this solution especially appropriate for this research. Moreover, they complement each other since pupillometry tracks rapid changes in cognitive load (high temporal resolution), while fNIRS can identify, with relatively good spatial resolution, which cortical areas become active during effortful listening.

The suitability of this approach was deemed appropriate for the laboratory experiment where we successfully recorded simultaneous pupil and cortical responses from groups of NH and CI listeners. Our results corroborated what previous research has suggested (Zekveld et al. 2014), that both techniques are likely to measure the same cognitive listening load. This was evident when looking at task-evoked physiological responses given that both pupil metrics (PPD and MPD), and LIFG amplitudes followed the same pattern of activation across conditions in both groups of participants. In the current experimental design, task-evoked physiological measures were not able to provide the complete picture of participants’ LE. However, thanks to pupillometry metrics (particularly the pupil baseline), we learnt about the increased arousal experienced by CI users during continuous background noise exposure. The simultaneous acquisition of both physiological measures was key to help us understand the results, preventing us from missing the potential underlying cognitive load of CI users that could not be registered by fNIRS results. The concurrent use of both techniques is thus recommended as a viable research tool for investigating the LE of people with and without HI in ecologically relevant listening environments.

The use of this approach however requires certain considerations. Care should be taken when placing the equipment on participants’ head. Gentle manipulation is required when fitting both the fNIRS headset and the pupil eye-tracker to avoid, first, the displacement of the CI magnet, and second, the repeated contact between the equipment and the CI microphone (behind the pinna). Moreover, keeping constant low luminance levels in the room is a requirement to prevent interference with both the fNIRS and the eye-tracker infrared lights.

Although the combination of both techniques has been proved feasible for research purposes, this approach may not be easily applicable in clinical settings where it is difficult to achieve a controlled environment for data acquisition. Nonetheless, the findings provided by this tool can certainly have clinical implications. A neural marker for real-world LE could help to identify those patients who are most at risk of experiencing excessive LE. Furthermore, this knowledge may inform the design of new interventions such as improved hearing devices and rehabilitation strategies to alleviate the mental effort and improve the lives of patients.

6.3.3 Listening efficiency as a novel outcome measure for CI users.

We used a joint analysis of accuracy and response time to obtain an integrated metric of the behavioural responses recorded in various studies (the laboratory experiment and the online test). Such a metric is the rate of evidence accumulation towards a correct response in a LBA model, and is interpreted as listening efficiency. Conceptually, it describes the amount of accuracy achieved per unit of effort expended. Therefore, listening efficiency is an objective outcome that reflects both intelligibility and LE. The advantages of such integrated measure are numerous. First, it combines the most important aspects (speech understanding and effort) that assess the effectiveness of HL treatments (hearing devices, listening training, etc.). Second, it provides better test-retest reliability (Bakun Emesh et al. 2021) than the analysis of accuracy and response time separately, and it characterises the SATO that is inherent to decision-making (Stafford et al. 2020). Third, the LBA analysis, using Bayes approach, considers the entire response time distribution, and thus uses information beyond just the mean or median. Moreover, the listening efficiency posterior distribution, being a probability function, ensures that the true value will lie within the HDI (95% most credible values). Finally, the practicality of this measure (based on behavioural responses) makes its use appropriate not only for research but also for audiological contexts.

Results from both studies confirmed our initial hypothesis that listening efficiency was sensitive to differences in cognitive load between groups of participants (between-subjects) and across conditions (within-subjects). Regarding the latter, listening efficiency was able to capture changes in task demands even when intelligibility remained almost perfect. Furthermore, listening efficiency was associated with different measures of LE in the patient group. Moderate correlations were found between CI users’ listening efficiency and their cognitive scores, subjective ratings, and physiological reactions. These correlations suggest that most

LE measures can potentially act as individual predictors of listening efficiency. Nonetheless, these associations were only found in the patient group, except for the pupil SL to baseline that was also predictive of listening efficiency in the NH group. The fact that these associations were mostly present in the CI group may be because LE is more relevant and more frequently experienced in the CI population.

In any case, the lack of correlations between measures of LE reported by many studies in the literature (Alhanbali et al. 2019; Anderson Gosselin and Gagné 2011; Hornsby 2013; McGarrigle, Rakusen, et al. 2021; Shields et al. 2023; Strand et al. 2018; Wendt et al. 2016; Zekveld and Kramer 2014) differ from our results. Perhaps the proposed listening efficiency metric, being an integration of intelligibility and effort, can assess more broadly participants' listening experience. As a consequence, it may partially tap into different underlying domains of the LE construct (e.g., cognitive, perceived, or exerted effort) that other measures seem to evaluate separately (Francis and Love 2019). Nonetheless, we have no evidence that the reason of finding correlations in this project as opposed to previous research is solely due to the integrated measure employed. Other reasons such as the experimental design and the plausible correlation analysis were also discussed in CHAPTER 2 as potential contributors that may have improved the consistency across measures.

Finally, since the listening efficiency measure is based on behavioural responses, it can be applied to both research and clinical contexts, where listening tasks and speech in noise tests are usually conducted. This is true even for listening tests that are performed online (as evidenced in CHAPTER 5), which expands even further the applicability of this measure. Nonetheless, its use in clinical environments would require the development of custom software to aid a prompt analysis and interpretation of results.

We argue that listening efficiency is a conceptually well-motivated metric that is easy to measure in practice and able to reflect both participants' speech understanding and effort during listening tasks. The fact that this metric captures small changes in cognitive load even when intelligibility remains high makes it suitable for being developed into a CI outcome measure. It holds promise to support the development and evaluation of a new breed of hearing technologies that aim to alleviate suprathreshold listening difficulties. Further research is needed to explore the sensitivity of this metric and the consistency with which it associates with other measures of LE under different experimental conditions.

6.3.4 Listening through a CI may be inherently effortful

The findings of all studies included in this thesis provided clear evidence that CI users experienced high level of LE in various communication scenarios commonly found in everyday life. These results, obtained from 168 CI recipients recruited throughout the project, were confirmed not only by self-reported but also by objective measures of listening efficiency and pupillometry metrics.

The laboratory experiment revealed that CI listeners were disproportionally affected by moderate ecologically relevant levels of background noise. CI participants exhibited significantly inferior listening efficiency than their NH peers when they listened to speech in a “cafeteria” background noise environment. Their physiological reactions also reflected this. Elevated levels of arousal as shown by larger pupil dilations during the background noise exposure suggested that CI users, unlike controls, may have been in a state of increased alertness to support their performance. Although the influence of other factors (including affective and emotional states, salience and acoustic characteristics of the sound, etc.) cannot be ruled out, such a physiological reaction is likely to reflect the cognitive processing and selective attention (top-down mechanism) that CI participants needed to identify any salient speech from the auditory background noise. This interpretation is consistent with participants’ behavioural and subjective responses (simultaneously recorded) during the task, and in line with previous research that showed (using vocoders) that the process of degraded speech is attention dependent (Wijayasiri et al. 2017; Wild et al. 2012).

Moreover, CI users needed longer to recover (manifested by prolonged dilations to baseline) from the arousal experienced during active listening compared to NH controls. Such levels of cognitive load were present even under highly favourable SNR (+20 dB) and when their intelligibility remained near ceiling. Thus, it is not surprising that the mental effort that CI users already experience in relatively easy listening conditions, will keep increasing as a function of task demands until reaching their cognitive capacity. Certainly, our results showed that the louder the background noise was, the larger their pupils became. This is suggestive of an increased allocation of cognitive resources needed to cope with the added difficulty. Although no signs of disengagement were observed during the experiment, one could assume that this sustained mental exertion if experienced on a daily basis may result in cognitive overload and potentially task withdrawal. In fact, CI users are considered to be in high risk of social isolation and loneliness (Mick et al. 2014; Pronk et al. 2011; Shukla et al. 2020).

Unfortunately, such a risk seemed to worsen during the COVID-19 pandemic as a results of the restrictions introduced to control the spread of the virus (Tagupa 2020). The survey results confirmed this. CI users reported actively avoiding some everyday communication scenarios due to the listening difficulties experienced. These difficulties included the perception of reduced speech intelligibility and increased LE both during in-person and remote communication. Overall, CI listeners who completed the survey struggled when having daily life conversations with people wearing facemasks or standing at a social distance. Likewise, the increased reliance on remote communication was perceived as additional communication challenge by our participants. Although these were subjective opinions that could have been influenced by other psychological factors such as general anxiety prevalent during the COVID-19 pandemic, we were able to corroborate some of these results from an objective point of view.

The online test provided evidence of the difficulties reported by CI users during online communication. Indeed, their listening efficiency when performing a behavioural test of speech recognition was significantly below that achieved by NH participants in all presentations modes (audio only, audio-visual, and audio-visual with captions). These results showed that online communication is more cognitively demanding for CI than for NH listeners, manifested as slower and less accurate responses to what they have heard. Yet, most CI users tended to persevere once they engage in an interaction as suggested by the low levels of disengagement reported throughout the project. However, the endurance of this cognitive load for sustained periods could affect their quality of life (McRackan, Hand, Velozo, et al. 2019; Skidmore et al. 2020), as well as their professional (Blumenthal and Sefotho 2022; Hetu et al. 1988; van der Hoek-Snieders et al. 2020; Kramer et al. 2006) and social interactions (Mick et al. 2014; Shukla et al. 2020).

Overall, these studies highlight the cognitive burden of listening through a CI in different real-world listening scenarios. By providing empirical evidence, this findings supports the notion that CI users experience high levels of LE in daily life, as previous research has suggested (Alhanbali et al. 2017).

6.3.5 Visual cues support CI users' listening efficiency in everyday life.

It is already known that visual information optimises CI recipients' speech recognition in real-world listening situations (Moberly et al. 2020). Results from our studies further confirmed the importance of visual cues for CI users during both in-person and remote communication. Most participants who completed the online survey reported relying significantly on lipreading and facial expressions to support their everyday conversations. Hence, listening situations where visual cues were not available were considered quite challenging and even avoided at times. Some of these situations involved listening to someone using a facemask or facial covering, through telephone, and through video call when the video camera and/or live captions were not available.

Regarding online communication, we obtained objective confirmation of the benefits provided by the presence of video and live captions during online video calls. The work presented in CHAPTER 5 revealed that CI users' listening efficiency improved considerably when visual information was present in comparison with the audio only condition, and even further with the addition of captions. The more visual cues were available, the more efficiently CI listeners processed and responded to the information presented. Therefore, although video call conversations continue to be significantly more demanding for CI users than for NH people, the presence of visual cues seemed to alleviate considerably the cognitive burden that they inherently experience during listening.

Other considerations were also made by CI recipients to enhance the advantages of visual cues. For instance, a slower speech pace was considered beneficial to facilitate verbal communication and lipreading for both in-person and remote listening scenarios. Likewise, according to CI listeners' preferences, transcriptions could be improved by showing bigger readable letters, different colours for the identification of different speakers, and positioned at mid-screen to avoid the extra effort of gaze switching between text and the speakers' faces.

Although visual cues were the most beneficial solution to support CI users' listening efficiency, the context also seemed to play a role in facilitating participants' intelligibility. It was mentioned in the survey that having previous knowledge of the topic to be discussed in a video call helped CI listeners to make sense of what they heard during the actual conversation. The online test results also verified this experience. Participants were more efficient at identifying a word heard when they correctly understood the first part of the sentence, which probably informed about the speech topic. In the same way, CI listeners seemed to rely on the context for the deduction of misheard words, given that semantic errors were made more often than other types of errors. However, the latter was a mere observation that did not reach statistical significance and therefore no conclusions could be drawn. It would be interesting to further explore what type of errors are more likely made by CI users during online communication.

6.4 LIMITATIONS

Although care was taken to maximise the quality of the research, the work presented here has several limitations to consider that are detailed below.

Recruitment into the three studies that constitute this project was conducted primarily through The NIHR Nottingham BRC participant database, and through national and regional hearing charities and organisations in the United Kingdom. In total 168 CI users and 75 age-matched NH controls participated in the studies.

Despite all efforts made, our sample was unbalanced in terms of gender and age, with most participants being females, in their seventies, and retired. We are aware that this profile may not necessarily represent the wider population of CI recipients. In addition, the fact that most volunteers were older may have contributed to the high levels of LE obtained throughout the project in the patient group. Indeed, it is known from previous research that aging generally leads to an increase in listening cognitive load (McGarrigle, Knight, et al. 2021; Phillips 2016; Tun et al. 2009). Nonetheless, the recruitment of an age-matched control group of NH participants aimed to counteract this limitation, so that differences in LE must be due to hearing rather than age-related causes. Moreover, the age distribution of the patient group seem to be consistent with UK surgical registration data that reported many adult CI users being aged 60 to 69 at the time of implantation (Raine 2014).

As explained briefly in section 3.6.1 (CHAPTER 3), one of the possible limitations of the laboratory experiment was the sample size of CI users. A sample of N=24 was considered adequate to obtain group level fNIRS data with good-to-excellent reliability, according to a fNIRS test-retest reliability study conducted previously in our laboratory (Wiggins et al. 2016). However, the high variability in speech understanding performance exhibited by our CI participants could also have been reflected in a diversity of brain activation. This would have prevented us from finding common cortical activity patterns among the implanted volunteers, and then contributed to the lack of statistical differences in the brain activity between both groups of participants. Such explanation might be feasible given that previous research has shown that CI users' cortical activation patterns correlate with their speech perception performance (Olds et al. 2016). Nonetheless, the present work cannot confirm it since our fNIRS analysis was limited to the group level, preventing us from drawing conclusions about single-subject level responses.

Likewise, the experimental design constrained the fNIRS brain imaging results to task-evoked responses. Such restriction impeded the assessment of LIFG activity at the noise baseline, when the main differences in cognitive load between groups might have been, as the pupil results suggested. This could have contributed to the failure in finding neural activity associated to LE. Moreover, based on previous research we directly targeted fronto-temporal regions, particularly the LIFG and RDLPFC, as a logical starting point for investigating effortful listening. However, there are other brain areas not covered in the study that are also believed to be associated with cognitive effort, such as the cingulo-operacular system and fronto-parietal regions (Eckert et al. 2016). The cortical activity of some of these areas, specifically the ones located at the outer cortex (e.g., anterior temporal-parietal areas), could certainly be assessed using fNIRS (Lawrence et al. 2018; White and Langdon 2021). With these limitations in place and without the entire picture of CI users' brain activity, we cannot discard the existence of a neural marker of LE in the implanted population although our results failed to identify it.

Throughout the project, we designed studies that aimed to reproduce listening scenarios commonly occurring in real life. Since we could not cover all listening situations that CI users encounter in daily life, we particularly focused on those that involved social interactions. For instance, listening to speech in a cafeteria background noise was specifically chosen to be representative of social conversations taking place in public settings such as cafes, restaurants, etc. Likewise, video call conversations were considered, as they became an increasingly frequent way of communication during the pandemic and beyond. To mimic these situations, we carefully selected stimuli that were as realistic as possible, e.g., meaningful sentences masked by realistic background noise at most frequent SNR levels. Nonetheless, there were aspects of the behavioural tasks that could not reproduced entirely a real listening experience. The first study, for instance, was restricted to a laboratory setup, and thus, it could not emulate the sound immersion typical of social environments. Instead, both the speech stimuli and the background noise were displayed from one single loudspeaker placed in front of the participants. Likewise, auditory speech was presented in isolation without showing the speaker's facial expressions. These limitations could have made the task to be perceived as more demanding than real life one-to-one conversations in social settings, making results not entirely representative of real life. Nonetheless, there exists some social situations (e.g., group conversations) where the listeners are not always able to identify and look towards the speaker at all times. Moreover, the correlations found between participants' listening efficiency during the task and the effort that they reported in daily life (measured by the EAS) suggest that the experimental task achieved at least modest ecological validity.

A similar restriction occurred with the online video call study that could not allow the usual interactions of real online conversations. Such flexibility is not possible in a controlled listening task where response time and accuracy measures are recorded. Nonetheless, the use of high context video clips from the BBC, covering a wide range of topics and presented by different speakers aimed to compensate for this limitation. In addition, the captions provided during the task were completely accurate, which again is not usually the case of many close captioning services that achieve inferior levels of accuracy. This may have increased the benefits provided by the captions, and thus, the listening efficiency of CI users in the "video with caption" condition. To this regard, it would be beneficial for any future work to investigate whether captions' accuracy may affect the listening efficiency of CI users during online video calls.

Another general constraint of the online studies compared to the laboratory experiment is the limited control over the testing conditions. For instance, it was not possible to check the sound reproduction settings of participants' computer during the video call study. Although we provided detailed set up instructions and asked participants to report which settings they were using, by no means we could know whether participants complied with the provided directives. Whilst this non-controlled testing environment entails an important limitation, at the same time, it conferred more realism to the task. In terms of ecological validity, it is positive that participants completed the test at home using the exact same devices and settings that they would normally use when having online video calls.

Finally, although most studies presented here aimed to use objective measures to assess the LE of CI listeners, such an approach was not possible in the online survey (CHAPTER 4). In-person laboratory work was cancelled or greatly restricted during various stages of the COVID-19 pandemic, which impeded us from continuing the same investigation line. Therefore, we decided to learn about the listening difficulties reported by the HL community during this new pressing situation. We only addressed CI listeners because they were the focus of this thesis and were affected greatly by the restrictions introduced to combat the spread of the virus (i.e., severe-to-profound). However, given the subjective nature of the survey, the results obtained were subject to recall bias. Moreover, although the survey was explicitly hearing focused, other psychological factors present during the COVID-19 pandemic such as general stress, and health anxiety might have influenced their responses. In addition, the lack of data from a control group prevented us from making claims about how specific our findings were to the CI-using population. Possibly, other collectives or even everyone was experiencing similar listening difficulties during the pandemic. Some of these limitations were partly amended by the online test (CHAPTER 5) that provided objective evidence about the difficulties that most CI listeners reported during online communication in comparison to NH controls.

6.5 IMPACT AND FUTURE DIRECTIONS

This project has provided insights about the LE of CI users in common real-world listening environments, both in-person and remotely. The findings based on multimodal measures of LE, including self-reported, behavioural, and physiological assessments, suggested that listening through a CI may be intrinsically effortful.

In particular, objective assessments were carried out using two innovative approaches: i) the simultaneously acquisition of fNIRS brain imaging and pupillometry as a tool to measure physiological changes in cognitive load; and ii) the listening efficiency measure that integrates both intelligibility and LE.

The listening efficiency metric proposed in this project has not, to the authors' knowledge, been previously employed in hearing sciences, nor in the LE literature. The LBA model (Brown and Heathcote 2008) employed in this thesis is able to perform a joint analysis of participants' behavioural measures (speed and accuracy), taking into account the underlying ability-trait (SATO) inherent to decision making, while improving the statistical power and the test-retest reliability (Stafford et al. 2020). Such an approach offers an integrated measure, interpreted here as listening efficiency, which reflects both intelligibility and LE. Throughout the project, we explored the sensitivity of this measure in a laboratory experiment and in an online test. In both cases, listening efficiency was able to capture differences in cognitive load between groups of CI and NH participants (between-subjects) and across SNR conditions (within-subjects), even when intelligibility remained near ceiling. Moreover, at the individual level this metric was associated with cognitive, subjective, and physiological measures of LE.

In the light of these promising results, we contend that listening efficiency should be considered as a novel CI outcome measure that assesses objectively not only the speech understanding performance of CI users but also the mental effort that they exert in doing so. Such integrated assessment is particularly relevant for clinical environments where intelligibility is considered in isolation without taking into account the complaints of many CI users about the cognitive effort that they experience on a daily basis (Gosselin and Gagné 2010). In fact, to date there is not a standardised measure of LE readily available for clinical use and thus, a single metric capable of offering a joint evaluation of intelligibility and LE holds promise as a candidate. Moreover, the correlation between listening efficiency and self-reported measures of LE indicates that, despite being an objective metric, it is in line with the cognitive challenges perceived by patients in everyday life. Therefore, its application in clinical practice can provide a more comprehensive and realistic assessment of patients' listening performance. Such an evaluation, being more representative of patients' listening experience, can better inform their clinical care and help clinicians choose the treatments that best meet their needs (e.g., fitting strategies, listening and cognitive trainings). Moreover, the listening efficiency measure can easily be obtained with the test batteries currently used by audiologists (Wilson et al. 2007). Nonetheless, its implementation in clinical settings would require further work aimed at reducing the processing time of the LBA analysis. It would require the development of custom software able to analyse the results and provide interpretable scores in a timely manner.

Future research should explore, in further detail, the sensitivity and practical utility of the listening efficiency metric, and the consistency in which it is associated with other measures of LE. This investigation can also be done by reanalysing the behavioural data from existing studies in the LE literature. Should listening efficiency be proved suitable, the ultimate goal should be its translation from research to clinical practice, so that it can be developed into a CI outcome measure.

The second approach, employed to investigate the neural correlates of LE, was the combination of two physiological measures: fNIRS-based brain imaging and pupillometry. Previous theories suggested that the LC-NE system, that is believed to be measured by pupillometry, in combination with frontal brain structures constitute a self-regulated system that optimise performance (Aston-Jones and Cohen 2005). Indeed, both pupillary responses and activity in some prefrontal areas, such as the LIFG, follow the same inverted U-shape pattern in relation to speech intelligibility (Aston-Jones and Cohen 2005; Lawrence et al. 2018; Poldrack et al. 2001; Zekveld et al. 2006). Greater responses occur at intermediate levels of task difficulty (degraded-yet-intelligible speech) when cognitive effort need to be spent to optimise performance, in comparison with both very easy (clear speech) and difficult (unintelligible noise) tasks that do not involve the recruitment of cognitive resources. Despite the close correspondence of both techniques at targeting LE, publications involving the simultaneous acquisition of functional brain imaging and pupillometry remain few (Zekveld et al. 2014). Hence, the novelty of using this approach in CI users.

Following the successful recording of fNIRS and pupil responses in NH and implanted adults, we can be further confident that the combination of these physiological measures is an apt and suitable research tool for the objective assessment of LE. Our findings showed that both techniques were able to reflect the same task-evoked cognitive load, as revealed by the same response patterns of PPD, MPD, and LIFG amplitude. Moreover, they seemed to complement each other, so that information not registered by fNIRS was provided by pupillometry. According to this, we believe that the concurrent use of both techniques can be a powerful research tool that provides a better understanding of the neural mechanism underlying LE in participants with different hearing profiles including CI users.

However, the project failed to identify a neural maker of LE in the implanted population. It remains unknown the reason of the similarity found in brain activity patterns between NH and CI participants. The limitations of the laboratory experiment previously described possibly prevented us from finding significant cortical activation associated with cognitive effort in CI users. Further research should be conducted to investigate whether LIFG activity may be elicited by the state of alertness that CI listeners exhibited (as revealed by pupil results at baseline) when listening to speech in the presence of realistic background noise. This could be done by adding a silent condition in a similar experimental paradigm, which would allow the recording of fNIRS responses in the contrast silence versus noise. Additionally, it would be interesting to learn how cortical patterns of brain activity in CI users may vary as a function of their individual differences in listening efficiency. In future investigations, the use of fNIRS and pupillometry might be able to provide insights into the neural correlates of real-world effortful listening in the implanted population.

The findings of this thesis have provided a better understanding of the cognitive challenges faced by CI users in daily life, and proposed two novel approaches to quantify objectively and non-invasively the cognitive demands of listening through a CI. Such assessments can be used in research and clinical contexts to allow the identification of those patients who are most at risk of experiencing excessive LE, so their clinical care could be modified accordingly. Moreover, they can inform the design of new interventions such as the development of improved hearing devices and rehabilitation strategies to alleviate the cognitive burden of listening with CI devices. Ultimately, this work would contribute to reducing LE, which is a key aspect of improving the lives of CI users, ensuring that they can communicate effectively, participate fully, and protect themselves against the long-term cognitive decline associated with HL.

6.6 CONCLUSIONS

Effortful listening is commonly experienced by CI listeners in daily life (Alhanbali et al. 2017). The findings presented in this thesis provides objective evidence that supports this notion. Despite good intelligibility, CI users showed higher levels of LE and reduced listening efficiency compared with NH controls in different real-world listening scenarios. Particularly, when listening to speech in social environments under moderate levels of background noise, when attending online video calls, and generally, when visual cues were not available to support speech comprehension. Although CI listeners tend to persevere once engaged in an interaction, this elevated mental exertion if sustained over time may lead to cognitive overload and social withdrawal. The objective assessments employed throughout the project warrant further consideration for both research and clinical purposes. The simultaneous acquisition of fNIRS-based brain imaging and pupillometry has proved a valuable approach for investigating the neural correlates of LE. Likewise, the listening efficiency measure has demonstrated sensitivity to differences in task demands and between groups, even when intelligibility remained near perfect. This metric was, at the individual level, associated with cognitive, subjective, and physiological measures of LE, all of which could act as individual predictors. We argue that listening efficiency, being able to reflect both intelligibility and effort, is a practical and suitable measure to quantify listening performance in realistic acoustic conditions. It thus holds promise for being developed into a CI outcome measure. Future research should further explore the sensitivity and practical utility of this metric across diverse listening situations.

APPENDICES

APPENDIX A: SURVEY ITEMS (CHAPTER 4)

Demographic Information

Q1. How old are you?

- *18-29 years*
- *30-39 years*
- *40-49 years*
- *50-59 years*
- *60-69 years*
- *70-79 years*
- *Over 80 years*

Q2. Which gender best defines you?

- *Female*
- *Male*
- *Other*
- *Prefer not to answer*

Q3. What is your highest level of education?

- *No formal schooling completed*
- *High school (i.e GCSEs or equivalent)*
- *Vocational Training (i.e apprentice)*
- *Further education (i.e A-levels or equivalent)*
- *Higher Education (i.e undergraduate university degree or equivalent)*
- *Postgraduate education (i.e masters degree or PhD)*
- *Unsure*
- *Prefer not to say*

Q4. What is your current employment status?

- *Homemaker*
- *Employed full time*
- *Employed part time*
- *Self-employed*
- *Student*
- *Unable to work*
- *Retired*
- *Furloughed or unable to work due to COVID-19*
- *Unemployed*
- *Unsure*
- *Other*

- *Prefer not to say*

Q5. What is your country of residence? Please select the country you are currently living in.

- *196 countries included in alphabetic order from Afghanistan to Zimbabwe.*

Hearing Information

Q6. For how long have you been using your cochlear-implant(s)? *(If you received two implants at different times, please answer this question based on your first implant.)*

- *Less than 6 months*
- *6 to 12 months*
- *1 to 3 years*
- *4 to 6 years*
- *7 to 10 years*
- *11 to 20 years*
- *More than 20 years*

Q7. Do you know roughly how old you were when you lost your hearing in your implanted ear(s)? Please indicate the option that applies according to the age at which you lost your hearing.

- *Infant (0-1 year)*
- *Toddler (1-3 years)*
- *Pre-schooler (3-5 years)*
- *Childhood (6-12 years)*
- *Teenager (13-19 years)*
- *Twenties*
- *Thirties*
- *Forties*
- *Fifties*
- *Sixties*
- *Seventies or over*
- *I don't know*

Q8. How many cochlear implants do you have?

- *One (unilateral)*
- *Two (bilateral)*

Q8.1. *(If one)* How is your hearing in the non-implanted ear? *If you are unsure, please select the option that you feel best describes your hearing in the non-implanted ear.*

- *Good (No loss)*
- *Mild loss*
- *Moderate loss*
- *Severe loss*
- *Profound loss*
- *Total loss*

Q8.2. *(If one)* Do you use a hearing aid in the opposite ear to your cochlear-implant?

- *Yes*
- *No*

Q8.2.1. *(If yes)* How often do you usually use your hearing aid in daily life?

- *8+ hours a day*
- *4 to 8 hours a day*
- *1 to 4 hours a day*
- *Occasionally (more than 1 hour a week)*
- *Seldom (less than 1 hour a week)*
- *Never*

Q8.2.2. *(If yes)* For how long have you been using the hearing aid in the opposite ear to your cochlear-implant?

- *Less than 6 months*
- *6 to 12 months*
- *1 to 3 years*
- *4 to 6 years*
- *7 to 10 years*
- *11 to 20 years*
- *More than 20 years*

Q8.2.3. Do you use any special feature that makes the coordination between your hearing aid and your cochlear-implant easier? *(If yes)* Please give more details.

Q9. In everyday life, to what extent do you rely on these ways of communication? *Scale: [“Never”, “Rarely”, “Sometimes”, “Often”, “Almost Always”].*

- *Listening*
- *Lipreading*

- *Facial expressions*
- *Sign Language*
- *Text transcriptions*

The use of facemasks

Q10. Have you been in a situation during which you had to talk with someone who was wearing a face covering or mask (e.g., in a shop, medical facility, etc.)?

Q10.1. *(If yes)* For each question below, please select the option that best reflects your experience in this or these situation(s). *Scale [0-4].*

-How much of the person's speech were you able to understand? *Anchors: "None" and "All"*

-How much mental effort did you have to put in to achieve this level of understanding? *Anchors: "No effort" and "Lots of effort"*

-How often did you ask the speaker to repeat (part of) the message? *Anchors: "Never" and "Almost Always"*

-How often did you give up trying to communicate because the effort required was too great? *Anchors: "Never" and "Almost Always"*

-Did you experience any feelings of anxiety or stress as a result of difficulty communicating? *Anchors: "Not at all" and "A great deal"*

-To what extent did the communication leave you feeling tired/fatigued? *Anchors: "Not at all" and "A great deal"*

Q10.2 *(If yes)* Considering your listening experiences <u>before and after the covid-19 outbreak</u>, how much do you think your communications have changed due to the speaker wearing a face mask? *Scale: ["Much less", "Less", "No difference", "More", "Much more"].*

-Due to the speaker's face covering, my understanding is:

-Due to the speaker's face covering, the level of effort I experience is:

-Due to the speaker's face covering, the frequency with which I ask the speaker to repeat the message is:

-Due to the speaker's face covering, the frequency with which I give up communicating is:

-Due to the speaker's face covering, the level of anxiety or stress I feel is:

-Due to the speaker's face covering, the level of tiredness/fatigue I feel is:

Q11. The following is a list of potential challenges associated with listening to someone who is wearing a facemask. Please rate how relevant they are according to your experience. *Scale: ["Not relevant", "Slightly relevant", "Moderately relevant", "Very relevant", "Extremely relevant"].*

-Muffled sound
-The person's voice is quieter
-Inability to see lip movements
-Inability to read facial expressions

Q12. How often do you find yourself avoiding face-to-face communication because of difficulty hearing someone who is wearing a facemask or covering?

Communication avoidance because of face mask use. *Scale: ["Never", "Rarely", "Sometimes", "Often", "Almost Always"].*

2 Metre distancing Rule

Q13. Have you been in a situation(s) during which you had to talk with someone at 2 metre distance (e.g., friend, family, shop assistant, etc.)?

Q13.1. *(If yes)* For each question below, please select the option that best reflects your experience in this or these situation(s). *Scale [0-4].*

-How much of the person's speech were you able to understand? *Anchors: "None" and "All"*

-How much mental effort did you have to put in to achieve this level of understanding? *Anchors: "No effort" and "Lots of effort"*

-How often did you ask the speaker to repeat (part of) the message? *Anchors: "Never" and "Almost Always"*

-How often did you give up trying to communicate because the effort required was too great? *Anchors: "Never" and "Almost Always"*

-Did you experience any feelings of anxiety or stress as a result of difficulty communicating? *Anchors: "Not at all" and "A great deal"*

-To what extent did the communication leave you feeling tired/fatigued? *Anchors: "Not at all" and "A great deal".*

Q13.2. *(If yes)* Considering your listening experiences <u>before and after the covid-19 outbreak</u>, how much do you think your communications have changed due to having to keep 2 metre away from others? *Scale: ["Much less", "Less", "No difference", "More", "Much more"]*.

-When I listen at 2 metre distance, my understanding is:

-When I listen at 2 metre distance, the level of effort I experience is:

-When I listen at 2 metre distance, the frequency with which I ask the speaker to repeat the message is:

-When I listen at 2 metre distance, the frequency with which I give up communicating is:

-When I listen at 2 metre distance, the level of anxiety or stress I feel is:

-When I listen at 2 metre distance, the level of tiredness/fatigue I feel is:

Q14. The following is a list of potential challenges associated with listening to someone from 2 metre distance. Please rate how relevant they are according to your experience: *Scale: [“Not relevant”, “Slightly relevant”, “Moderately relevant”, “Very relevant”, “Extremely relevant”].*

-Background noise is more intrusive

-The speech sounds more echoey

-The person’s voice is quieter

-It’s harder to lipread from 2 metres away

Q15. How often do you find yourself avoiding face-to-face communication because of difficulty hearing someone who is 2 metres apart?

Communication avoidance because of 2 metre distance rule. *Scale: [“Never”, “Rarely”, “Sometimes”, “Often”, “Almost Always”].*

Q16. If you have any additional comments, please write them here: (e.g., comments about the previous measures mentioned or other physical measures that make communication difficult)

Telephone and video calls

Q17.How often do you have telephone or video calls....? *Scale: [“Never”, “Rarely”, “Sometimes”, “Often”, “Almost Always”].*

...to communicate with family and friends?

...to access essential services (such as health and care consultations, shops, pharmacy, etc.)?

...for work or other commitments?

Q18. How has the frequency of telephone and video calls changed since the covid-19 outbreak?

Telephone and video call frequency since the covid-19 outbreak. *Scale: ["Much less", "Less", "No difference", "More", "Much more"]*.

Telephone calls

Q19. Do you have conversations on the telephone (such as on a mobile or a landline)?

Q19.1. *(If yes)* For each question below, please select the option that best reflects your experience when you have a (mobile or landline) telephone call. *Scale [0-4].*

-How much of the person's speech are you able to understand? *Anchors: "None" and "All"*

-How much mental effort do you have to put in to achieve this level of understanding? *Anchors: "No effort" and "Lots of effort".*

-How often do you ask the speaker to repeat (part of) the message? *Anchors: "Never" and "Almost Always".*

-How often do you give up trying to communicate because the effort required was too great? *Anchors: "Never" and "Almost Always".*

-Do you experience any feelings of anxiety or stress as a result of difficulty communicating by telephone? *Anchors: "Not at all" and "A great deal".*

-What is the level of tiredness/fatigue you usually feel after having a telephone call? *Anchors: "Not at all" and "A great deal".*

Q19.2. *(If yes)* The following is a list of potential challenges associated with telephone conversations. Please rate how relevant they are according to your experience. *Scale: ["Not relevant", "Slightly relevant", "Moderately relevant", "Very relevant", "Extremely relevant"].*

-Background noise in your environment

-Volume too low

-Poor quality telephone line

-Unfamiliar voice or accent

-Fast speech pace

Q20. How often do you find yourself avoiding telephone calls because of difficulty understanding what is being said?

Telephone call avoidance. *Scale: [“Never”, “Rarely”, “Sometimes”, “Often”, “Almost Always”].*

Video calls

Q21. Do you communicate using video calls or video conferencing? (for example using Skype, Zoom, Teams, FaceTime, etc.)

Q21.1. *(If yes)* For each question below, please select the option that best reflects your experience when you have a video call. *Scale [0-4].*

-How much of the person’s speech are you able to understand? *Anchors: “None” and “All”*

-How much mental effort do you have to put in to achieve this level of understanding? *Anchors: “No effort” and “Lots of effort”.*

-How often do you ask the speaker to repeat (part of) the message? *Anchors: “Never” and “Almost Always”.*

-How often do you give up trying to communicate because the effort required was too great? *Anchors: “Never” and “Almost Always”.*

-Do you experience any feelings of anxiety or stress as a result of difficulty communicating by video? *Anchors: “Not at all” and “A great deal”.*

-What is the level of tiredness/fatigue you usually feel after having a video call? *Anchors: "Not at all" and "A great deal".*

Q21.2. *(If yes)* The following is a list of potential challenges associated with video calls/conferences. Please rate how relevant they are according to your experience: *Scale: ["Not relevant", "Slightly relevant", "Moderately relevant", "Very relevant", "Extremely relevant"].*

-Background noise in your environment
-Volume too low
-Unclear who was speaking (group video calls)
-Too many people speaking at the same time
-Audio and video out of sync.
-Audio or video cutting in and out (connection problems)
-Poor or lack of transcriptions
-Speaker's video camera turned off
-Unfamiliar voice or accent
-Fast speech pace

Q22. How often do you find yourself avoiding video calls because of difficulty understanding what is being said?

Video call avoidance. *Scale: ["Never", "Rarely", "Sometimes", "Often", "Almost Always"].*

Solutions to minimise impact

Q23. To what extent do you think these solutions could help to improve your everyday life communications? If any of these solutions doesn't apply to you, please check the "Not applicable" option. *Scale: ["Not at all", "A little", "Somewhat", "A lot", "A great deal"].*

-More access to face-to-face services

-The use of transparent face masks

-Reduced background noise in public places

-Use of speech-to-text (live transcription) apps

-Stream sounds from phone and video call directly to hearing devices

-Improved confidence to use video calling

-Increased volume in phone and video calls

-Slower speech pace

-Real-time transcriptions during video calls

-Video call recording apps (allowing re-watching of the call afterwards)

-Improved bandwidth during video calls (less cutting out of audio or video)

-Making sure that the person speaking always has their camera turned on during video calls

Q24. Are there any other solutions you have tried or are thinking of trying to improve your everyday life communications?

APPENDIX B: THEMES AND CATEGORIES IDENTIFIED IN THE ONLINE SURVEY'S TEXT RESPONSES (CHAPTER 4).

Table B. 1. *Themes, categories, number of mentions, and example statements in response to question Q8.2.3. "Do you use any special feature that makes the coordination between your hearing aid and your cochlear-implant easier? (If yes) Please give more details".*

Theme	Category	Number	Example
Special feature to coordinate hearing devices	Contralateral routing of signals solution	10	"Have a Naida link which allows sounds to be transferred from CI to HA, it enables me to hear environmental sounds that I wouldn't normally pick up in my non implant ear"
	External microphone	1	"I wear an AB CI and Phonak hearing aid that dual access my Rogers Pen"

Table B. 2. *Themes, categories, number of mentions, and example statements in response to question Q16. "If you have any additional comments, please write them here: (e.g., comments about the previous measures mentioned or other physical measures that make communication difficult)".*

Theme	Category	Number	Example
Facemasks	Major problem	17	"Face mask coverings make life very hard", "The voice muffling is the most difficult thing"
	In medical settings	5	"I had to go for a CT scan and the nurses insisted on keeping their masks on. Very stressful and I think they skimped on what info they gave me because of the difficulties"
	People collaboration	5	"Thankfully, people are following U.K. guidance and removing their mask if they need to speak to me"
	No major impact	3	"I have not had major problems understanding people wearing face masks. I do need to slightly increase the number of tomes I ask them to repeat, but that is a consequence of face masks for mostly everyone"
	Low confidence/ loneliness	3	"Avoid going anywhere where you have to wear a mask. Lose my confidence"
	Clear visor	1	"Clear visor masks are helpful, face coverings impossible-terrible situation for deaf people"
Social distancing	No major impact unless noisy environment	7	"I don't find the 2m rule a problem unless is a noisy environment"
	Difficult to lip-read	4	"Being outside at 2 metres apart just doesn't work for using Lipreading to help with communication. Many times I have told the person I am with that I cannot keep 2 metres"
Combination of facemask and social distancing	Impossible to communicate	2	"Combination of 2 metre distance PLUS face mask = nightmare!"
Shields	Muffled sounds	2	"Shields at counters, especially those made up of strips cause issues due to muffled sound and light glare"
Listening training	Helpful	2	"I spend hours to train what I hear and less dependence on other form of communication which allow me to feel more confidence in these times"
Other feelings/ attitudes	Lack of people collaboration	3	"people give up"
	Perseverance	1	"My nature is not to give up"

Table B. 3. *Themes, categories, number of mentions (N.), and example statements in response to question Q24. “Are there any other solutions you have tried or are thinking of trying to improve your everyday life communications?”*

Theme	Category	N.	Example
Text transcriptions/ live subtitles	Essential for video calls but not always available	8	“Video calls with subtitles would make life easier for me”, “Zoom should provide free captions like Google does”
	Text transcription apps	6	“When I join a Zoom meeting on my laptop I have the app "Live Transcribe" open on my mobile phone beside the loudspeaker. It's not perfect but fills in some gaps and gives me clues if I lose the thread. Also it is retained on the phone so I could go back and check if necessary provided it is still there (limited time recorded)”
	Not always accurate	1	“I use live captions in video meetings but they aren't very accurate and once told me people were talking about thin crispy zombies!”
Stream sounds to hearing devices	Improve speech clarity	6	“I have purchased a USB headset that streams direct to my cochlear implants via my Roger Select and an adapter. I take numerous Skype calls and meetings daily and can hear almost every word. It has been a life saver”
	Overlapping talk	1	“Other people I struggle to hear at all even with the streamer. When people start talking over each other I tend to give up!”
Telephone and video call Avoidance	Too challenging	5	“I don't use Skype, or the telephone too challenging”
	Use of text messaging and other services	2	“I use RelayUK to make calls and sometimes receive them. Texts and emails are my lifeblood!”
Transparent mask/face visors	Allow Lipreading	5	“I wear a visor not a mask,I find pointing to it and saying that I lipread is an instant,constant reminder to people that I have hearing needs. I would quite like to see visors with "please speak clearly!" printed across the headband!
External microphone	Useful for social distancing	3	“I can pass my MiniMic to an individual to use when speaking to me from 2m away as will pick up on Bluetooth that way”
	Covid transmission risk	1	“Have been unable to use Mini-mike for speakers because of Covid transmission risk.”
Ask people collaboration	Remove facemasks	3	“Being very specific saying I cannot Lipread with a mask and I need them to remove it when talking to me. Sometimes it works!”
	Help from family/friends	2	“My wife is my hearing support. If she were not here my life would be very very different!”
Speakers’ camera during video calls	Camera on/ correct placement	2	“Camera placement to see speaker faces is a big problem. I hate seeing just top of head. Speaker cannot see themselves in little windows and mostly cannot when using mobile devices”
Better bandwidth	Quality of Audio/ Video & Audio in sync	2	“On video more bandwidth to audio to give much better lower frequency transmission would be fantastic”
Previous Knowledge	Topic and people in the conversation	2	“Having a good idea of the subject-matter the other person seeks to talk about enables one to better 'select' the vocabulary base and, thereby, the sense and meaning of what they are saying”
Other solutions	over-the-ear headphones	1	“I wear over the ear headphones for telephone and video calls”
	Perseverance	1	“There is little more I can do but I persevere as much as possible.”
	Wear a badge	1	“I wear a badge stating that I am deaf and that has helped as people are aware of my problem”
	Reduce background noise	1	“I'm also much more bothered by background noise than I was before, so being able to cut out background would be really helpful.”
	Meditation	1	“Meditation, to accept and get used to the ‘new normal’.”

APPENDIX C: TEST ITEMS (CHAPTER 5)

C.1 Demographic and hearing questions

Demographic Information

Q1. How old are you? Please, enter your age in years into the box below:

Q2. Which gender best defines you?

- *Female*
- *Male*
- *Other*
- *Prefer not to answer*

Q3. What is your current employment status?

- *Homemaker*
- *Employed full time*
- *Employed part time*
- *Self-employed*
- *Student*
- *Unable to work*
- *Retired*
- *Furloughed or unable to work due to COVID-19*
- *Unemployed*
- *Unsure*
- *Other*
- *Prefer not to say*

Hearing Information

The following questions are about your hearing in the left ear.

Q4. How is your hearing in your left ear? *If you are unsure, please select the option that you feel best describes your hearing in this ear.*

- *Good- no loss*
- *mild loss*
- *moderate loss*
- *severe loss*
- *profound loss*

Q5. Do you use a hearing device in your left ear?

- *None*
- *hearing aid*
- *cochlear implant*
- *other*

If cochlear implant selected:

Q5.1. For how long have you been using your cochlear- implant in the left ear? Please enter a number in years into the box below. *Enter 0 if you have less than one year experience with your implant.*

Q5.2. Do you know roughly how old you were when you lost your hearing in your left ear? *Please indicate the option that applies according to the age at which you lost your hearing.*

- *Infant (0-1 year)*
- *Toddler (1-3 years)*
- *Pre-schooler (3-5 years)*
- *Childhood (6-12 years)*
- *Teenager (13-19 years)*
- *Twenties*
- *Thirties*
- *Forties*
- *Fifties*
- *Sixties*
- *Seventies or over*
- *I don’t know*

If hearing aid selected:

Q5.1. Do you use your hearing aid regularly on a daily basis?

- *Yes*
- *No*

Q5.2. Are you using the hearing aid right now (during the experiment)?

- *Yes*
- *No*

If other device selected:

Q5.1. If you selected other please write the name of the device you use in your left ear in the box below:

The following questions are about your hearing in the right ear.

Q6. How is your hearing in your right ear? *If you are unsure, please select the option that you feel best describes your hearing in this ear.*

- *Good- no loss*
- *mild loss*
- *moderate loss*
- *severe loss*
- *profound loss*

Q7. Do you use a hearing device in your right ear?

- *None*
- *hearing aid*
- *cochlear implant*
- *other*

If cochlear implant selected:

Q7.1. For how long have you been using your cochlear- implant in the right ear? Please enter a number in years into the box below. *Enter 0 if you have less than one year experience with your implant.*

Q7.2. Do you know roughly how old you were when you lost your hearing in your right ear? *Please indicate the option that applies according to the age at which you lost your hearing.*

- *Infant (0-1 year)*
- *Toddler (1-3 years)*
- *Pre-schooler (3-5 years)*
- *Childhood (6-12 years)*
- *Teenager (13-19 years)*
- *Twenties*

- *Thirties*
- *Forties*
- *Fifties*
- *Sixties*
- *Seventies or over*
- *I don't know*

If hearing aid selected:

Q7.1. Do you use your hearing aid regularly on a daily basis?

- *Yes*
- *No*

Q7.2. Are you using the hearing aid right now (during the experiment)?

- *Yes*
- *No*

If other device selected:

Q7.1. If you selected other please write the name of the device you use in your right ear in the box below:

Ways of communication

Q8. In everyday life, to what extent do you rely on these ways of communication?

Please provide an answer to each way of communication mentioned below.

	Never	Rarely	Sometimes	Often	Almost Always
Listening					
Lipreading					
Facial expressions					
Sign Language					
Text transcriptions					

C.2 Setup instructions

Getting set up

Please make sure you wear your hearing devices e.g., cochlear implant(s) or hearing aid(s) as you would normally do. Once adjusted, please don´t do any further adjustment during the experiment.

This study requires you to watch and listen to video and audio clips. You can use either your computer/tablet loudspeakers or headphones. Once you decide how you will listen, please use the same setup throughout the experiment.

Next button

Q9. Which sound reproduction setting are you using during the experiment? *Please use the set up that you would normally use during a video call or the one you feel more comfortable with.*

- *Loudspeakers from computer or tablet*
- *Headphones*
- *Stream sounds to your hearing devices*
- *Other*

Next button

Sound check

1.Click the play button below to hear an audio example.

2.Adjust the volume on your device so that the sound is at a comfortable level (not too quiet, not too loud). You can replay it as many times as you like.

If you didn´t hear anything, ensure your sound is not muted and that the volume is turned up sufficiently, then press play again.

Play button

☐ By clicking the box I confirm that the sound is at a comfortable level.

Button Selection

After listening to the audio and video clips, you will be required to complete a task by selecting some buttons. If you have a touchscreen computer, you can tap the buttons with your finger. Otherwise, you can use your mouse to click the buttons. Whatever option you decide, please use the same approach throughout the experiment.

Q10. I confirm that I am using a:

- *Touchscreen device (e.g., tablet)*
- *Computer with mouse*

Next button

Great! Now you are ready to start the test!

Please, do not make any changes to your "device settings, change the volume or switch to any other apps" until the experiment is complete.

If you are ready to start, click 'Continue'

To go through the instructions again, click 'Back'

Back button ***Continue button***

APPENDIX D: THEMES AND CATEGORIES IDENTIFIED IN THE ONLINE TEST’S TEXT RESPONSES (CHAPTER 5)

D.1 COCHLEAR IMPLANT GROUP

Conversation A

Table D. 1. *Themes, categories, number of mentions (N.), and example statements given by CI users in Conversation A, in response to questions: “if this was a real video call that you were involved in, which mode would you prefer to use?”, “please tell us why this was your preferred mode”, “were there any downsides to your preferred mode?”. The number in brackets beside each presentation mode indicates the number of participants who selected that particular mode.*

Mode	Theme	Category	N.	Example
Audio (1)	Reason	Difficult lipreading	1	“I preferred the audio, as with the video I found the males lips hard to read”
	Downside	Accent	1	“The accent”
Video (3)	Reason	Lipreading	2	“I could read their lips which aided understanding whereas audio only I found the participants speech a little fast to follow at first”
		Facial expressions	1	“See faces”
		Caption distracting/inaccurate	1	“I prefer it to the mode with the closed captions as this is sometimes distracting as they load one word at a time. Furthermore, the captions are often slightly inaccurate and this can make my understanding worse as I am receiving inputs that don't match”
	Downside	Connection quality	1	“Relying on video depends on the internet connection. If the connection is bad the video can be blurry which makes it hard to lipread. A bad connection can also distort the sound”
		Video quality/camera position	1	“Shadow on the ladies face and sitting to the side slightly”
		Caption needed depending on conversation type	1	“if a technical conversation or background noise would use captioning too”
Captions (46)	Reason	Lipreading not possible (unfamiliar accents/unclear speech, audio-video not in sync)	17	“The man was quite difficult to follow, even with lip reading. He didn't move his mouth in a natural manner”
		Allows to understand all conversation/ less effortful	15	“With the captions I could understand all the conversation.”, “Captions - I found it was less effort to understand the conversation.”
		Provides confirmation/confidence	10	“the captions help to confirm my understanding and clarify any missed words”, “It gave me confidence”.
		Accurate/ in-sync transcriptions	3	“The text seemed very accurate although it didn't filter out the speakers' hesitations and repetitions.”,” The captions were in sinc with the speakers in this clip”
		Useful when poor sound quality/noisy background	3	“there was some interference, a whistling sound that detracted from the audio”
		Useful when poor video quality/ obscured faces	3	“captions essential as video quality not good”,”The female participan't face was partly

				shadowed. When faces are obscured or blurred on a video call it makes things that bit harder”
		Useful when Unfamiliar topic/ difficult conversation	3	“The context was a lot harder to grasp - maybe that made it harder too”
		Useful when overlapping speech	2	“they talked over each other… so captions were helpful”
		Useful for fast speech	1	“The speed of the talking and lack of lipmovement from the participants left me baffled”
	Downside	No downsides (if accurate, readable, and in sync with speech)	21	“No, I followed it okay”, “No, as long as the captions are big enough to read easily and keep pace with the dialogue.”
		not accurate/delayed	6	“captions was slow to catch up”, “if it is automatic captioning, it often is wrong in accuracy”.
		Need to concentrate on text and miss people’s faces	5	“I miss some of the interaction by looking down at the captions”
		Cognitive overload when looking at captions and people at the same time	3	“Cognitive overload as usual... if I'm looking at the text then I'm not looking at the faces. Perhaps, in a more ideal world, each person's speech would appear in speech bubbles besides their heads... like a comic book”
		Absent during overlapping speech	2	“Lots of talking over each other in that video so no captions at times”
		Subtitle position	1	“The positioning of the subtitles, it would be useful to have some control over that”
		Difficulty to know whom the caption refer to	1	“One is not always sure to whom the captions refer”

Conversation B

Table D. 2. *Themes, categories, number of mentions (N.), and example statements given by CI users in Conversation B, in response to questions: "if this was a real video call that you were involved in, which mode would you prefer to use?", "please tell us why this was your preferred mode", "were there any downsides to your preferred mode?". The number in brackets beside each presentation mode indicates the number of participants who selected that particular mode.*

Mode	Theme	Category	N.	Example
Audio (2)	Reason	Clear speech/Easy topic	2	"Participants speech was clear and topic easily followed so didn't need video"
	Downside	None	2	"None at all"
Video (13)	Reason	Clear speech	7	"very clearly enunciated and easily followed"
		Facial expressions/body language	6	"I like to see the faces and expressions"
		Easy to lipread	4	"Video mode allowed me to lip read"
		Captions distracting (out of sync/delayed)	3	"I found audio only hard to hear and video with captions distracting from speech and hard to follow."
	Downside	None	10	"Not really"
		Video and audio out of sync	1	"The delay in video and speech was distracting"
		Concentration required	1	"Requires concentration to follow the conversation"
		No captions as back-up	1	"If I missed the odd word there was no way of checking (i.e. no captions)"
Captions (35)	Reason	Allows to understand all conversation/less effortful	17	"made it so much easier to understand with the captions and less effort involved with straining to lipread and make use of residual hearing"
		Provides confirmation/ back-up/confidence	12	"I barely looked at the captions, but it was good to have them there. To quickly cast my eye down to see what that missed word was", " I can relax more knowing that the captions are there so that I don't miss out."
		Useful when audio and video out of sync	2	"because the talking was delayed in what was being said"
		Useful for fast speech/volume variations	2	"because the speak too fast for me and their pitch is too high"
		Useful when lipreading is difficult (accents)	2	"I found the female with the glasses hard to lip read"
	Downside	None	19	"Not really. it would be fine to participate in a real conversation in this mode."
		Miss people's faces	3	"by looking down at the captions I miss a bit of the visual interaction and facial expressions"
		Cognitive overload	1	"Cognitive overload... particularly with trying to synchronise the speech and captions in my head. In everyday life, I would only look at the text when I'm unsure of what's been said."
		Captions out of sync (Delayed)	4	"Captions don't always sinc with lip patterns"
		not accurate	1	"auto palantypist doesn't always get it right"
		Lack of punctuation	1	"Punctuation the captions would improve things..."
		Difficulty to know whom the caption refer to	2	"It would be better if different peoples conversations were in different coloured text, as its difficult to understand who is saying what!"
		Overlapping talk	2	"people tend to talk across each other"
		Captions position	1	"Yes, the biggest problem is where the subtitles are situated on the screen. They are usually down the bottom of the screen which makes it difficult to interact with the individual, I would prefer them to be mid screen"

Conversation C

***Table D. 3**. Themes, categories, number of mentions, and example statements given by CI users in Conversation C, in response to questions: "if this was a real video call that you were involved in, which mode would you prefer to use?", "please tell us why this was your preferred mode", "were there any downsides to your preferred mode?". The number in brackets beside each presentation mode indicates the number of participants who selected that particular mode.*

Mode	Theme	Category	N.	Example
Video (19)	Reason	Clear speech/good sound quality	10	"Both speakers were easy to understand and I had no problem following the conversation. Even when the captions were showing, I didn't need use them"
		Easy to lipread (video and audio in sync)	8	"the lip-sync was accurate so the combination of lip-reading and sound made it easy to understand"
		Captions out of sync	7	"I would normally want the captions, but they are out of sync with the speaker, so I have gone for the video."
		Seeing the speakers improves the conversation experience (facial expressions/body language)	5	"I was able to understand both people very easily and clearly but the video added to the experience by showing their facial expressions and obvious enjoyment of the conversation. I did not need to use subtitles."
		Makes conversation more personal	1	"Just more personal to have a video call"
	Downside	None	10	"None; in this case the video was good and the voices and lips were synchronised"
		No captions	6	"I can't check from the captions when I'm not certain what was said."
		Video and audio out of sync	2	"In some parts there was a lag between the video & audio that could have made lipreading difficult"
		Concentration required	1	"I have to concentrate to understand but I am still able to follow the conversation well"
Captions (31)	Reason	Provides confirmation/back up	13	"It helps to be able to double check what I thought I heard was correct"
		Allows to understand all conversation/less effortful	13	"I could follow every word with the video with captions without concentration", "Those speaking in the videos, were speaking exceptionally clearly in ideal conditions, with not background noises. Watching speakers, lipreading plus captioning is the best accessibility practice."
		Useful when lipreading is difficult (accents, distance from camera, fast speech)	4	"the pattern of the lips was sometimes difficult to follow", "It's really hard work to concentrate to lipread people 'further away' than if it were a face to face conversation. Also people speaking too fast and I cannot ask them to slow down a bit"
		Accustomed to watching TV with subtitles	2	"I have become used to watching the television with subtitles in fact if there was a programme I wanted to watch and there were no subtitles I tend to abandon it."
	Downside	None	19	"Not as long as the captions are easy to see and keep pace with the dialogue."
		Not in sync	5	"The syncing of text to voice is a little off, so there's a feeling of cognitive overload"
		Not accurate	3	"some of the words were wrong"
		Miss people's faces	3	"With captions I could not look at the faces but that is a small price to pay"
		Requires ability to read quickly	1	"One has to read very quickly."

D.2 NORMAL HEARING GROUP

Conversation A

Table D. 4. *Themes, categories, number of mentions (N.), and example statements given by NH participants in Conversation A, in response to questions: "if this was a real video call that you were involved in, which mode would you prefer to use?", "please tell us why this was your preferred mode", "were there any downsides to your preferred mode?". The number in brackets beside each presentation mode indicates the number of participants who selected that particular mode.*

Mode	Theme	Category	N.	Example
Audio (6)	Reason	Clear/ Easier to follow	3	"The dialogue is clear and I am able to focus on the conversation. If the audio was unclear I would opt for visual and then text , if really bad"
		No distraction from video or captions	3	"The images were a little distracting as one of the presenters was not sitting straight in front of the screen"
	Downside	None	4	"No"
		No facial expressions	1	"If I were talking to family or business I would then prefer the visual to see facial expressions"
		Long pauses	1	"There were quite long pauses so potentially you could be wondering if the connection was lost without visual clues"
Video (32)	Reason	Facial expressions/see who is talking	16	"could put a face to the voice and know which of them was talking"
		Easier to understand (if poor audio quality or overlapping talk)	8	"Could understand better what was being said especially when they talked at same time"
		Caption distracting/inaccurate	7	"Captions were sometimes inaccurate and generally distracting."
		Feels more natural/ better engaged/inclusive	6	"Again, it felt natural. I felt better engaged.", "more like F2F encounters".
		More common in daily meetings	2	"I use this mode in daily work life"
	Downside	None	19	"None in this mode"
		Video can be distracting	4	"Shadow on one person slightly annoying", "Only the distraction of unfamiliar backgrounds"
		Difficult to understand when overlapping talk	3	"Still difficult to understand when both participants were speaking but captions did not resolve that problem either in this example."
		Captions useful for accents, unfamiliar or unclear speech	3	"I would certainly appreciate captions if a person's accent or dialect was less familiar and/or difficult for me to understand quickly"
		Connection quality/background noise	2	"Slight buzz noise in background "
		Cognitive load	1	"Video calls are more tiring than audio calls"
Captions (12)	Reason	Useful when poor sound quality, noisy background or unclear speech	7	"There was a buzzing sound in the background which was a little off putting and the audio had a slight echo in the video option, so I preferred to read the captions in this instance for extra reassurance"
		Useful when overlapping talk	4	"At some points the speakers spoke at the same time and it was difficult to make out what they were saying so it was useful to have the captions."
		Easier to understand/more clarity	3	"I can see who is talking and clarity on the words spoken, if the transcription is accurate"
	Downside	Miss visual information/facial expressions	5	"I miss some of the interaction by looking down at the captions"

		Not accurate	4	“Inaccuracies in auto subtitling”
		None	3	“No, I could see the participants, hear them and read the transcript as a backup.”
		Difficult to read due to lack of punctuation and scrolling text	2	“no punctuation and the transcript disappears after a couple of lines”

Conversation B

Table D. 5*. Themes, categories, number of mentions (N.), and example statements given by NH participants in Conversation B, in response to questions: “if this was a real video call that you were involved in, which mode would you prefer to use?”, “please tell us why this was your preferred mode”, “were there any downsides to your preferred mode?”. The number in brackets beside each presentation mode indicates the number of participants who selected that particular mode.*

Mode	**Theme**	**Category**	**N.**	**Example**
Audio (4)	Reason	Clear speech	2	“The sound and voices were very clear and there was no need to see any of the presenters on a screen to keep up with the conversation”
		Only one input to concentrate on	2	“With audio there was only one thing to concentrate on”
	Downside	None	2	“None at all”
		Requires good hearing and concentration	2	“Relies on good hearing and concentration”
		Lose the personal interaction	1	“Lose the personal touch and ability to see facial expression of those speaking”
Video (40)	Reason	Facial expressions enhance comprehension (who speaks, their attitude, the tone of the conversation)	29	“I can see who is speaking as well their facial expressions which helps to understand the tone of the conversation”, “Prefer to see facial expression as this adds meaning to the words”
		It’s more pleasant, more personal and feels more involved and natural	11	“felt more involved, more natural, real people”, “felt more like a normal face to face conversation.”
		Captions distracting or out of sync	7	“Captions were distracting and unnecessary”, ”captions delayed”
		Lipreading	3	“Partial lipreading”,” “I also use lip-reading even if people speak clearly and I do not have hearing loss”
	Downside	None	29	“Not that I can think of”
		Video can be distracting	5	“Possibly could get distracted by things that could be seen other than conversation”
		Captions needed for missing words (especially when accents, unfamiliar speech or technical discussion)	5	“no, but if someone had a strong accent or English was not their first language, I might also read the live transcription”
		Video and audio out of sync	2	“Audio-Visual lag”
		Poor video quality	1	“Quality of picture”
Captions (6)	Reason	Easier to understand/more clarity	5	“More clarity”, “Sometimes you can't make out exactly what a person said and the captions are a good fail safe.”
		Useful when poor connection signal	1	“Best proof against disturbances in the signal”
	Downside	None	5	“No. Text helped supplement”
		Miss visual information	1	“You might miss some visual information while looking at the scrolling text.”

Conversation C

Table D. 6. *Themes, categories, number of mentions (N.), and example statements given by NH participants in Conversation C, in response to questions: "if this was a real video call that you were involved in, which mode would you prefer to use?", "please tell us why this was your preferred mode", "were there any downsides to your preferred mode?". The number in brackets beside each presentation mode indicates the number of participants who selected that particular mode.*

Mode	Theme	Category	N.	Example
Audio (5)	Reason	Clear speech	2	"clear speakers, no interruptions, so concentrated on content"
		Audio and video out of sync	3	"The sound wasn't quite in sync with the video which I found a little off putting. The text caption wasn't fully synched either so I preferred to just listen to this one"
	Downside	None	2	"No, I can listen and do other tasks at the same time"
		Miss facial expressions	3	"I missed seeing the lively expressive faces"
Video (36)	Reason	Facial expressions enhance comprehension (who speaks, their attitude, the tone of the conversation)	22	"I wanted to see the expressions on their faces", "I can observe expressions and see that someone is not distracted or disengaged"
		Captions distracting (not accurate, lack of capitalization, out of sync)	12	"I do find having text can be distracting", "I found the transcription irritating as there were lower case letters where there should have been capitals (I am a grammar geek!)"
		It's more pleasant, more personal and feels more engaging and natural	10	"I like to see the people I am speaking to, particularly as we've not had much opportunity for interaction for 2 years."
		Lipreading	3	"I also use lip-reading even if people speak clearly and I do not have hearing loss"
		Common mode in daily meetings	2	"It's my usual way of holding a meeting over the internet"
	Downside	None	22	"None I can think of"
		Captions useful if missing words (especially when accents, unclear or fast speech)	7	"Delivery a mixture of slow and fast sections of speech so text may assist understanding"
		Video and audio out of sync	3	"At some points in the video there was a mismatch between the video and the sound"
		Video can be distracting	3	"some backgrounds can be distracting"
		Be on camera	1	"I have to be on camera as well which means I have to be careful with my reactions."
Captions (9)	Reason	Easier to understand/more clarity (especially when accents, unclear or fast speech)	6	"it helped to understand better what they were saying.", "Sometimes you may mishear or misinterpret what was said because of regional accents"
		Useful when poor connection signal (audio and video out of sync)	3	"It was easier to follow the conversation as the sound and vision wasn't always in sync'"
		Help to stay focus/remember the conversation	2	"They were speaking quite slowly so it was hard to stay focussed. The captions to read made it more interesting"
	Downside	None	4	"No"
		Miss facial expressions	3	"Couldn't really read the captions and see the expressions at the same time"
		Not accurate	2	"Just the questionable reliability of auto subtitling."

REFERENCES

Ackerman, P. L., and American Psychological Association. 2011. *Cognitive Fatigue: Multidisciplinary Perspectives on Current Research and Future Applications*. American Psychological Association.

Adank, Patti. 2012. 'The Neural Bases of Difficult Speech Comprehension and Speech Production: Two Activation Likelihood Estimation (ALE) Meta-Analyses'. *Brain and Language* 122(1):42–54. doi: 10.1016/j.bandl.2012.04.014.

Afouras, Triantafyllos, Joon Son Chung, Andrew Senior, Oriol Vinyals, and Andrew Zisserman. 2019. 'Deep Audio-Visual Speech Recognition'. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1–1. doi: 10.1109/TPAMI.2018.2889052.

Ahern, Sylvia, and Jackson Beatty. 1979. 'Pupillary Responses During Information Processing Vary with Scholastic Aptitude Test Scores'. *Science* 205(4412):1289–92.

Akeroyd, Michael A. 2008. 'Are Individual Differences in Speech Reception Related to Individual Differences in Cognitive Ability? A Survey of Twenty Experimental Studies with Normal and Hearing-Impaired Adults.' *International Journal of Audiology* 47 Suppl 2:S53-71. doi: 10.1080/14992020802301142.

Alhanbali, Sara, Piers Dawes, Simon Lloyd, and Kevin J. Munro. 2017. 'Self-Reported Listening-Related Effort and Fatigue in Hearing-Impaired Adults'. *Ear and Hearing* 38(1):e39–48. doi: 10.1097/AUD.0000000000000361.

Alhanbali, Sara, Piers Dawes, Rebecca E. Millman, and Kevin J. Munro. 2019. 'Measures of Listening Effort Are Multidimensional': *Ear and Hearing* 40(5):1084–97. doi: 10.1097/AUD.0000000000000697.

Alhanbali, Sara, Kevin J. Munro, Piers Dawes, Peter J. Carolan, and Rebecca E. Millman. 2020. 'Dimensions of Self-Reported Listening Effort and Fatigue on a Digits-in-Noise Task, and Association with Baseline Pupil Size and Performance Accuracy'. *International Journal of Audiology* 1–11. doi: 10.1080/14992027.2020.1853262.

Anderson, Carly A., Ian M. Wiggins, Pádraig T. Kitterick, and Douglas E. H. Hartley. 2019. 'Pre-Operative Brain Imaging Using Functional Near-Infrared Spectroscopy Helps Predict Cochlear Implant Outcome in Deaf Adults'. *Journal of the Association for Research in Otolaryngology*. doi: 10.1007/s10162-019-00729-z.

Anderson Gosselin, Penny, and Jean-Pierre Gagné. 2011. 'Older Adults Expend More Listening Effort Than Young Adults Recognizing Speech in Noise'. *Journal of Speech, Language, and Hearing Research* 54(3):944–58. doi: 10.1044/1092-4388(2010/10-0069).

Annis, Jeffrey, Brent J. Miller, and Thomas J. Palmeri. 2017. 'Bayesian Inference with Stan: A Tutorial on Adding Custom Distributions'. *Behavior Research Methods* 49(3):863–86. doi: 10.3758/s13428-016-0746-9.

Annis, Jeffrey, and Thomas J. Palmeri. 2018. 'Bayesian Statistical Approaches to Evaluating Cognitive Models'. *WIREs Cognitive Science* 9(2):31. doi: https://doi.org/10.1002/wcs.1458.

ASHA. 2020. ‘Hearing Loss and Remote Work: Advice for Effective Communication During Virtual Meetings’. *American Speech-Language-Hearing Association*. Retrieved 25 January 2022 (https://www.asha.org/public/hearing-loss-and-remote-work-advice-for-effective-communication-during-virtual-meetings/).

Aspinall, Angie. 2021. ‘Online Meetings Are Easier for People with Hearing Loss’. *Hearing Like Me*. Retrieved 25 January 2022 (https://www.hearinglikeme.com/online-meetings-are-easier-for-people-with-hearing-loss/).

Aston-Jones, Gary, and Jonathan D. Cohen. 2005. ‘AN INTEGRATIVE THEORY OF LOCUS COERULEUS-NOREPINEPHRINE FUNCTION: Adaptive Gain and Optimal Performance’. *Annual Review of Neuroscience* 28(1):403–50. doi: 10.1146/annurev.neuro.28.061604.135709.

Atcherson, Samuel R., Lisa Lucks Mendel, Wesley J. Baltimore, Chhayakanta Patro, Sungmin Lee, Monique Pousson, and M. Joshua Spann. 2017. ‘The Effect of Conventional and Transparent Surgical Masks on Speech Understanding in Individuals with and without Hearing Loss’. *Journal of the American Academy of Audiology* 28(01):058–067. doi: 10.3766/jaaa.15151.

Azadarmaki, Roya, Rhonda Tubbs, Douglas A. Chen, and Frank G. Shellock. 2014. ‘MRI Information for Commonly Used Otologic Implants: Review and Update’. *Otolaryngology–Head and Neck Surgery* 150(4):512–19. doi: 10.1177/0194599813518306.

Babiloni, Claudio, Vittorio Pizzella, Cosimo Del Gratta, Antonio Ferretti, and Gian Luca Romani. 2009. ‘Fundamentals of Electroencefalography, Magnetoencefalography, and Functional Magnetic Resonance Imaging.’ *International Review of Neurobiology* 86:67–80. doi: 10.1016/S0074-7742(09)86005-4.

Badajoz-Davila, Javier, Jörg M. Buchholz, and Richard Van-Hoesel. 2020. ‘Effect of Noise and Reverberation on Speech Intelligibility for Cochlear Implant Recipients in Realistic Sound Environments’. *The Journal of the Acoustical Society of America* 147(5):3538–49. doi: 10.1121/10.0001259.

Baddeley, Alan. 2012. ‘Working Memory: Theories, Models, and Controversies’. *Annual Review of Psychology* 63(1):1–29. doi: 10.1146/annurev-psych-120710-100422.

Bakun Emesh, Tamar, Dror Garbi, Alon Kaplan, Hila Zelicha, Anat Yaskolka Meir, Gal Tsaban, Ehud Rinott, and Nachshon Meiran. 2021. ‘Retest Reliability of Integrated Speed–Accuracy Measures’. *Assessment* 1073191120985609. doi: 10.1177/1073191120985609.

Baltes, Paul B., and Ulman Lindenberger. 1997. ‘Emergence of a Powerful Connection Between Sensory and Cognitive Functions Across the Adult Life Span: A New Window to the Study of Cognitive Aging?’ *Psychology and Aging* 12(1):12–21. doi: https://doi.org/10.1037/0882-7974.12.1.12.

Barker, Alex B., Paul Leighton, and Melanie A. Ferguson. 2017. ‘Coping Together with Hearing Loss: A Qualitative Meta-Synthesis of the Psychosocial Experiences of People with Hearing Loss and Their Communication Partners’. *International Journal of Audiology* 56(5):297–305. doi: 10.1080/14992027.2017.1286695.

Başkent, Deniz. 2012. ‘Effect of Speech Degradation on Top-Down Repair: Phonemic Restoration with Simulations of Cochlear Implants and Combined Electric–Acoustic Stimulation’. *Journal of the Association for Research in Otolaryngology* 13(5):683–92. doi: 10.1007/s10162-012-0334-3.

Basura, Gregory J., Xiao-Su Hu, Juan San Juan, Anne-Michelle Tessier, and Ioulia Kovelman. 2018. ‘Human Central Auditory Plasticity: A Review of Functional near-Infrared Spectroscopy (FNIRS) to Measure Cochlear Implant Performance and Tinnitus Perception: FNIRS and Human Central Auditory Plasticity’. *Laryngoscope Investigative Otolaryngology* 3(6):463–72. doi: 10.1002/lio2.185.

Beatty, Jackson. 1982. ‘Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources.’ *Psychological Bulletin* 91(2):276–92. doi: 10.1037/0033-2909.91.2.276.

Benarroch, Eduardo E. 2018. ‘Locus Coeruleus’. *Cell and Tissue Research* 373(1):221–32. doi: 10.1007/s00441-017-2649-1.

Bench, John, Åse Kowal, and John Bamford. 1979. ‘The Bkb (Bamford-Kowal-Bench) Sentence Lists for Partially-Hearing Children’. *British Journal of Audiology* 13(3):108–12. doi: 10.3109/03005367909078884.

Benjamini, Yoav, and Yosef Hochberg. 1995. ‘Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing’. *Journal of the Royal Statistical Society. Series B (Methodological)* 57(1):289–300.

Bennett, Andrew A., Emily D. Campion, Kathleen R. Keeler, and Sheila K. Keener. 2021. ‘Videoconference Fatigue? Exploring Changes in Fatigue after Videoconference Meetings during COVID-19.’ *Journal of Applied Psychology* 106(3):330–44. doi: 10.1037/apl0000906.

Bernstein, Lynne E., Edward T. Auer, and Sumiko Takayanagi. 2004. ‘Auditory Speech Detection in Noise Enhanced by Lipreading’. *Speech Communication* 44(1–4):5–18. doi: 10.1016/j.specom.2004.10.011.

Besser, Jana, Thomas Koelewijn, Adriana A. Zekveld, Sophia E. Kramer, and Joost M. Festen. 2013. ‘How Linguistic Closure and Verbal Working Memory Relate to Speech Recognition in Noise—A Review’. *Trends in Amplification* 17(2):75–93. doi: 10.1177/1084713813495459.

Bhatti, Anam, Hamza Akram, Hafiz Basit, Ahmed Khan, Syeda Mahwish, Raza Naqvi, and Muhammad Bilal. 2020. ‘E-Commerce Trends during COVID-19 Pandemic’. *International Journal of Future Generation Communication and Networking* 13.

Bianchi, Federica, Sébastien Santurette, Dorothea Wendt, and Torsten Dau. 2016. ‘Pitch Discrimination in Musicians and Non-Musicians: Effects of Harmonic Resolvability and Processing Effort’. *Journal of the Association for Research in Otolaryngology* 17(1):69–79. doi: 10.1007/s10162-015-0548-2.

Binder, Jeffrey R., Einat Liebenthal, Edward T. Possing, David A. Medler, and B. Douglas Ward. 2004. ‘Neural Correlates of Sensory and Decision Processes in Auditory Object Identification’. *Nature Neuroscience* 7(3):295–301. doi: 10.1038/nn1198.

Blumenthal, Lior, and Maximus M. Sefotho. 2022. ‘The Effects of Cognitive Effort on Academic Performance of Learners with Cochlear Implants in a Private Mainstream School in Gauteng’. *African Journal of Disability* 11. doi: 10.4102/ajod.v11i0.886.

Boisvert, Isabelle, Mariana Reis, Agnes Au, Robert Cowan, and Richard C. Dowell. 2020. ‘Cochlear Implantation Outcomes in Adults: A Scoping Review’ edited by S. Usami. *PLOS ONE* 15(5):e0232421. doi: 10.1371/journal.pone.0232421.

Bokolo, Anthony Jnr. 2020. ‘Use of Telemedicine and Virtual Care for Remote Treatment in Response to COVID-19 Pandemic’. *Journal of Medical Systems* 44(7):132. doi: 10.1007/s10916-020-01596-5.

Boksem, Maarten A. S., and Mattie Tops. 2008. ‘Mental Fatigue: Costs and Benefits’. *Brain Research Reviews* 59(1):125–39. doi: 10.1016/j.brainresrev.2008.07.001.

Boucsein, Wolfram. 2012. *Electrodermal Activity*. Second Edition. Springer US.

Bradshaw, J. L. 1968. ‘Pupil Size and Problem Solving’. *Quarterly Journal of Experimental Psychology* 20(2):116–22. doi: 10.1080/14640746808400139.

Brehm, Jack W., and Elizabeth A. Self. 1989. ‘The Intensity of Motivation’. *Annual Review of Psychology* 40:109–31. doi: https://doi.org/10.1146/annurev.ps.40.020189.000545.

Brigadoi, Sabrina, Lisa Ceccherini, Simone Cutini, Fabio Scarpa, Pietro Scatturin, Juliette Selb, Louis Gagnon, David A. Boas, and Robert J. Cooper. 2014. ‘Motion Artifacts in Functional Near-Infrared Spectroscopy: A Comparison of Motion Correction Techniques Applied to Real Cognitive Data’. *NeuroImage* 85:181–91. doi: 10.1016/j.neuroimage.2013.04.082.

Brooks, Samantha K., Rebecca K. Webster, Louise E. Smith, Lisa Woodland, Simon Wessely, Neil Greenberg, and Gideon James Rubin. 2020. ‘The Psychological Impact of Quarantine and How to Reduce It: Rapid Review of the Evidence’. *The Lancet* 395:912–20. doi: https://doi.org/10.1016/S0140-6736(20)30460-8.

Brooks, Stephen P., and Andrew Gelman. 1998. ‘General Methods for Monitoring Convergence of Iterative Simulations’. *Journal of Computational and Graphical Statistics* 7(4):434–55. doi: 10.1080/10618600.1998.10474787.

Brown, Scott D., and Andrew Heathcote. 2008. ‘The Simplest Complete Model of Choice Response Time: Linear Ballistic Accumulation’. *Cognitive Psychology* 57(3):153–78. doi: 10.1016/j.cogpsych.2007.12.002.

BSA. 2019. *Assessment of Speech Understanding in Noise in Adults with Hearing Difficulties*. *British Society of Audiology Practice Guidance*. British Society of Audiology Practice Guidance.

Buchman, Craig A., René H. Gifford, David S. Haynes, Thomas Lenarz, Gerard O’Donoghue, Oliver Adunka, Allison Biever, Robert J. Briggs, Matthew L. Carlson, Pu Dai, Colin L. Driscoll, Howard W. Francis, Bruce J. Gantz, Richard K. Gurgel, Marlan R. Hansen, Meredith Holcomb, Eva Karltorp, Milind Kirtane, Jannine Larky, Emmanuel A. M. Mylanus, J. Thomas Roland, Shakeel R. Saeed, Henryk Skarzynski, Piotr H. Skarzynski, Mark Syms, Holly Teagle, Paul H. Van de Heyning, Christophe Vincent, Hao Wu, Tatsuya Yamasoba, and Terry Zwolan. 2020. ‘Unilateral Cochlear Implants for Severe, Profound, or Moderate Sloping to Profound Bilateral Sensorineural Hearing Loss: A Systematic Review and Consensus Statements’. *JAMA Otolaryngology–Head & Neck Surgery*. doi: 10.1001/jamaoto.2020.0998.

Buchman, Craig A., Jacques A. Herzog, Jonathan L. McJunkin, Cameron C. Wick, Nedim Durakovic, Jill B. Firszt, Dorina Kallogjeri, and CI532 Study Group. 2020. ‘Assessment of Speech Understanding After Cochlear Implantation in Adult Hearing Aid Users: A Nonrandomized Controlled Trial’. *JAMA Otolaryngology–Head & Neck Surgery* 146(10):916. doi: 10.1001/jamaoto.2020.1584.

Bürkner, Paul-Christian. 2017. ‘Brms: An R Package for Bayesian Multilevel Models Using Stan.’ *Journal of Statistical Software* 80(1):1–28. doi: 10.18637/jss.v080.i01.

Caldwell, Meredith T., Nicole T. Jiam, and Charles J. Limb. 2017. ‘Assessment and Improvement of Sound Quality in Cochlear Implant Users’. *Laryngoscope Investigative Otolaryngology* 2(3):119–24. doi: 10.1002/lio2.71.

Calhoun, V. D., M. C. Stevens, G. D. Pearlson, and K. A. Kiehl. 2004. ‘FMRI Analysis with the General Linear Model: Removal of Latency-Induced Amplitude Bias by Incorporation of Hemodynamic Derivative Terms’. *NeuroImage* 22(1):252–57. doi: 10.1016/j.neuroimage.2003.12.029.

Cannata, Davide, and Sam Redfern. 2020. ‘OpenFaceR: Developing an R Package for the Convenient Analysis of OpenFace Facial Information.’ *EasyChair Preprint* (4389):10.

Carlsson, Per-Inge, Jennie Hjaldahl, Anders Magnuson, Elisabeth Ternevall, Margareta Edén, Åsa Skagerstrand, and Radi Jönsson. 2015. ‘Severe to Profound Hearing Impairment: Quality of Life, Psychosocial Consequences and Audiological Rehabilitation’. *Disability and Rehabilitation* 37(20):1849–56. doi: 10.3109/09638288.2014.982833.

Cattell, Raymond B. 1966. ‘The Scree Test For The Number Of Factors’. *Multivariate Behavioral Research* 1(2):245–76. doi: 10.1207/s15327906mbr0102_10.

CHABA. 1977. *Speech Understanding and Aging*. Committee on Hearing, Bioacoustics, and Biomechanics. Washington, D.C.: National Academies Press. doi: 10.17226/20619.

Chia, Ee-Munn, Jie Jin Wang, Elena Rochtchina, Robert R. Cumming, Philip Newall, and Paul Mitchell. 2007. ‘Hearing Impairment and Health-Related Quality of Life: The Blue Mountains Hearing Study.’ *Ear and Hearing* 28(2):187–95. doi: 10.1097/AUD.0b013e31803126b6.

Chodosh, Joshua, Barbara E. Weinstein, and Jan Blustein. 2020. ‘Face Masks Can Be Devastating for People with Hearing Loss’. *BMJ* 370:m2683. doi: 10.1136/bmj.m2683.

Chung, Joon Son, Andrew Senior, Oriol Vinyals, and Andrew Zisserman. 2017. ‘Lip Reading Sentences in the Wild’. Pp. 3444–53 in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI: IEEE.

Cochrane, D., and G. H. Orcutt. 1949. ‘Application of Least Squares Regression to Relationships Containing Auto- Correlated Error Terms’. *Journal of the American Statistical Association* 44(245):32–61. doi: 10.2307/2280349.

Coez, Arnaud, Monica Zilbovicius, Evelyne Ferrary, Didier Bouccara, Isabelle Mosnier, Emmanuèle Ambert-Dahan, Eric Bizaguet, Jean-Luc Martinot, Yves Samson, and Olivier Sterkers. 2014. ‘Brain Voice Processing with Bilateral Cochlear Implants: A Positron Emission Tomography Study’. *European Archives of Oto-Rhino-Laryngology* 271(12):3187–93. doi: 10.1007/s00405-013-2810-8.

Cooper, Robert J., Juliette Selb, Louis Gagnon, Dorte Phillip, Henrik W. Schytz, Helle K. Iversen, Messoud Ashina, and David A. Boas. 2012. ‘A Systematic Comparison of Motion Artifact Correction Techniques for Functional Near-Infrared Spectroscopy’. *Frontiers in Neuroscience* 6. doi: 10.3389/fnins.2012.00147.

Corbetta, Maurizio, and Gordon L. Shulman. 2002. ‘Control of Goal-Directed and Stimulus-Driven Attention in the Brain’. *Nature Reviews Neuroscience* 3(3):201–15. doi: 10.1038/nrn755.

Crosson, Bruce, Anastasia Ford, Keith M. McGregor, Marcus Meinzer, Sergey Cheshkov, Xiufeng Li, Delaina Walker-Batson, and Richard W. Briggs. 2010. ‘Functional Imaging and Related Techniques: An Introduction for Rehabilitation Researchers’. *The Journal of Rehabilitation Research and Development* 47(2):vii. doi: 10.1682/JRRD.2010.02.0017.

Dalton, Dayna S., Karen J. Cruickshanks, Barbara E. K. Klein, Ronald Klein, Terry L. Wiley, and David M. Nondahl. 2003. ‘The Impact of Hearing Loss on Quality of Life in Older Adults.’ *The Gerontologist* 43(5):661–68. doi: 10.1093/geront/43.5.661.

Daneman, Meredyth, and Patricia A. Carpenter. 1980. ‘Individual Differences in Working Memory and Reading’. *Journal of Verbal Learning and Verbal Behavior* 19(4):450–66. doi: 10.1016/S0022-5371(80)90312-6.

Davis, Matthew H., and Ingrid S. Johnsrude. 2003. ‘Hierarchical Processing in Spoken Language Comprehension’. *The Journal of Neuroscience* 23(8):3423–31. doi: 10.1523/JNEUROSCI.23-08-03423.2003.

Deary, Ian J. 1994. ‘Intelligence and Auditory Discrimination: Separating Processing Speed and Fidelity of Stimulus Representation’. *Intelligence* 18(2):189–213. doi: 10.1016/0160-2896(94)90027-2.

Delpy, D. T., and M. Cope. 1997. ‘Quantification in Tissue Near-Infrared Spectroscopy’. *Philosophical Transactions of the Royal Society B: Biological Sciences* 352(1354):649–59. doi: 10.1098/rstb.1997.0046.

Devocht, Elke M. J., A. Miranda L. Janssen, Josef Chalupper, Robert J. Stokroos, and Erwin L. J. George. 2017. ‘The Benefits of Bimodal Aiding on Extended Dimensions of Speech Perception: Intelligibility, Listening Effort, and Sound Quality’. *Trends in Hearing* 21:233121651772790. doi: 10.1177/2331216517727900.

Dimitrijevic, Andrew, Michael L. Smith, Darren S. Kadis, and David R. Moore. 2019. ‘Neural Indices of Listening Effort in Noisy Environments’. *Scientific Reports* 9(1):11278. doi: 10.1038/s41598-019-47643-1.

Dincer D’Alessandro, Hilal, Deborah Ballantyne, Patrick J. Boyle, Elio De Seta, Marco DeVincentiis, and Patrizia Mancini. 2018. ‘Temporal Fine Structure Processing, Pitch, and Speech Perception in Adult Cochlear Implant Recipients': *Ear and Hearing* 39(4):679–86. doi: 10.1097/AUD.0000000000000525.

Dingemanse, J. Gertjan, and André Goedegebure. 2019. 'The Important Role of Contextual Information in Speech Perception in Cochlear Implant Users and Its Consequences in Speech Tests'. *Trends in Hearing* 23:233121651983867. doi: 10.1177/2331216519838672.

Dionisio, D. P., E. Granholm, W. A. Hillix, and W. F. Perrine. 2001. 'Differentiation of Deception Using Pupillary Responses as an Index of Cognitive Processing.' *Psychophysiology* 38(2):205–11.

Donkin, Chris, Lee Averell, Scott Brown, and Andrew Heathcote. 2009. 'Getting More from Accuracy and Response Time Data: Methods for Fitting the Linear Ballistic Accumulator'. *Behavior Research Methods* 41(4):1095–1110. doi: 10.3758/BRM.41.4.1095.

Dosenbach, Nico U. F., Kristina M. Visscher, Erica D. Palmer, Francis M. Miezin, Kristin K. Wenger, Hyunseon C. Kang, E. Darcy Burgund, Ansley L. Grimes, Bradley L. Schlaggar, and Steven E. Petersen. 2006. 'A Core System for the Implementation of Task Sets'. *Neuron* 50(5):799–812. doi: 10.1016/j.neuron.2006.04.031.

Dryden, Adam, Harriet A. Allen, Helen Henshaw, and Antje Heinrich. 2017. 'The Association Between Cognitive Performance and Speech-in-Noise Perception for Adult Listeners: A Systematic Literature Review and Meta-Analysis'. *Trends in Hearing* 21:233121651774467. doi: 10.1177/2331216517744675.

Duffy, Elizabeth. 1957. 'THE PSYCHOLOGICAL SIGNIFICANCE OF THE CONCEPT OF "AROUSAL" OR "ACTIVATION"'. *Psychol Rev.* 64(5):265–75. doi: 10.1037/h0048837.

Dunn, Camille C., Elizabeth Stangl, Jacob Oleson, Michelle Smith, Octav Chipara, and Yu-Hsiang Wu. 2020. 'The Influence of Forced Social Isolation on the Auditory Ecology and Psychosocial Functions of Listeners with Cochlear Implants During COVID-19 Mitigation Efforts'. *Ear & Hearing* Publish Ahead of Print. doi: 10.1097/AUD.0000000000000991.

Eckert, Mark A., Susan Teubner-Rhodes, and Kenneth I. Vaden. 2016. 'Is Listening in Noise Worth It? The Neurobiology of Speech Recognition in Challenging Listening Conditions'. *Ear Hear* 37:101S-110S. doi: 10.1097/AUD.0000000000000300.

Eisenberg, Laurie S., Amy Schaefer Martinez, Suzanne R. Holowecky, and Stephanie Pogorelsky. 2002. 'Recognition of Lexically Controlled Words and Sentences by Children with Normal Hearing and Children with Cochlear Implants': *Ear and Hearing* 23(5):450–62. doi: 10.1097/00003446-200210000-00007.

Elo, Satu, and Helvi Kyngäs. 2008. 'The Qualitative Content Analysis Process'. *Journal of Advanced Nursing* 62(1):107–15. doi: 10.1111/j.1365-2648.2007.04569.x.

Elwell, Clare E., and Chris E. Cooper. 2011. 'Making Light Work: Illuminating the Future of Biomedical Optics.' *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences* 369(1955):4358–79. doi: 10.1098/rsta.2011.0302.

Eurich, Bernhard, Thomas Klenzner, and Michael Oehler. 2019. ‘Impact of Room Acoustic Parameters on Speech and Music Perception among Participants with Cochlear Implants’. *Hearing Research* 377:122–32. doi: 10.1016/j.heares.2019.03.012.

Evans, Nathan J., and Eric-Jan Wagenmakers. 2020. ‘Evidence Accumulation Models: Current Limitations and Future Directions’. *The Quantitative Methods for Psychology* 16(2):73–90. doi: 10.20982/tqmp.16.2.p073.

Ferrari, Marco, and Valentina Quaresima. 2012. ‘A Brief Review on the History of Human Functional Near-Infrared Spectroscopy (FNIRS) Development and Fields of Application’. *NeuroImage* 63(2):921–35. doi: 10.1016/j.neuroimage.2012.03.049.

Festen, Joost M., and Reinier Plomp. 1990. ‘Effects of Fluctuating Noise and Interfering Speech on the Speech-reception Threshold for Impaired and Normal Hearing’. *The Journal of the Acoustical Society of America* 88(4):1725–36. doi: 10.1121/1.400247.

Finger, Holger, Caspar Goeke, Dorena Diekamp, Kai Standvoß, and Peer König. 2017. ‘LabVanced: A Unified JavaScript Framework for Online Studies’. in *Science*. Cologne, Germany.

Firszt, Jill B., Laura K. Holden, Margaret W. Skinner, Emily A. Tobey, Ann Peterson, Wolfgang Gaggl, Christina L. Runge-Samuelson, and P. Ashley Wackym. 2004. ‘Recognition of Speech Presented at Soft to Loud Levels by Adult Cochlear Implant Recipients of Three Cochlear Implant Systems’: *Ear and Hearing* 25(4):375–87. doi: 10.1097/01.AUD.0000134552.22205.EE.

Foo, Catharina, Mary Rudner, Jerker Ronnberg, and Thomas Lunner. 2007. ‘Recognition of Speech in Noise with New Hearing Instrument Compression Release Settings Requires Explicit Cognitive Storage and Processing Capacity.’ *Journal of the American Academy of Audiology* 18(7):618–31.

Forstmann, B. U., R. Ratcliff, and E. J. Wagenmakers. 2016. ‘Sequential Sampling Models in Cognitive Neuroscience: Advantages, Applications, and Extensions.’ *Annual Review of Psychology* 67:641–66. doi: 10.1146/annurev-psych-122414-033645.

Forstmann, B. U., M. Tittgemeyer, E. J. Wagenmakers, J. Derrfuss, D. Imperati, and S. Brown. 2011. ‘The Speed-Accuracy Tradeoff in the Elderly Brain: A Structural Model-Based Approach’. *Journal of Neuroscience* 31(47):17242–49. doi: 10.1523/JNEUROSCI.0309-11.2011.

Forstmann, Birte, Scott Brown, Gilles Dutilh, Jane Neumann, and Eric-Jan Wagenmakers. 2010. ‘The Neural Substrate of Prior Information in Perceptual Decision Making: A Model-Based Analysis’. *Frontiers in Human Neuroscience* 4.

Francis, Alexander L., and Jordan Love. 2019. ‘Listening Effort: Are We Measuring Cognition or Affect, or Both?’ *Wiley Interdisciplinary Reviews: Cognitive Science* e1514. doi: 10.1002/wcs.1514.

Francis, Alexander L., and Jordan Love. 2020. ‘Listening Effort: Are We Measuring Cognition or Affect, or Both?’ *WIREs Cognitive Science* 11(1). doi: 10.1002/wcs.1514.

Francis, Alexander L., Megan K. MacPherson, Bharath Chandrasekaran, and Ann M. Alvar. 2016. ‘Autonomic Nervous System Responses During Perception of Masked Speech

May Reflect Constructs Other than Subjective Listening Effort'. *Frontiers in Psychology* 7. doi: 10.3389/fpsyg.2016.00263.

Fu, Qian-Jie, and Geraldine Nogaki. 2005. 'Noise Susceptibility of Cochlear Implant Users: The Role of Spectral Resolution and Smearing'. *Journal of the Association for Research in Otolaryngology* 6(1):19–27. doi: 10.1007/s10162-004-5024-3.

Gatehouse, S., and J. Gordon. 1990. 'Response Times to Speech Stimuli as Measures of Benefit from Amplification.' *British Journal of Audiology* 24(1):63–68.

Gatehouse, Stuart, and William Noble. 2004. 'The Speech, Spatial and Qualities of Hearing Scale (SSQ)'. *International Journal of Audiology* 43(2):85–99. doi: 10.1080/14992020400050014.

Gaylor, James M., Gowri Raman, Mei Chung, Jounghee Lee, Madhumathi Rao, Joseph Lau, and Dennis S. Poe. 2013. 'Cochlear Implantation in Adults: A Systematic Review and Meta-Analysis'. *JAMA Otolaryngology–Head & Neck Surgery* 139(3):265. doi: 10.1001/jamaoto.2013.1744.

Gernsbacher, Morton Ann. 2015. 'Video Captions Benefit Everyone'. *Policy Insights from the Behavioral and Brain Sciences* 2(1):195–202. doi: 10.1177/2372732215602130.

Ghinea, G., and J. P. Thomas. 1998. 'QoS Impact on User Perception and Understanding of Multimedia Video Clips'. Pp. 49–54 in *Proceedings of the sixth ACM international conference on Multimedia - MULTIMEDIA '98*. Bristol, United Kingdom: ACM Press.

Gilley, Phillip M., Anu Sharma, and Michael F. Dorman. 2008. 'Cortical Reorganization in Children with Cochlear Implants.' *Brain Research* 1239:56–65. doi: 10.1016/j.brainres.2008.08.026.

Gilzenrat, Mark S., Sander Nieuwenhuis, Marieke Jepma, and Jonathan D. Cohen. 2010. 'Pupil Diameter Tracks Changes in Control State Predicted by the Adaptive Gain Theory of Locus Coeruleus Function.' *Cognitive, Affective & Behavioral Neuroscience* 10(2):252–69. doi: 10.3758/CABN.10.2.252.

Goldin, Alexander, Barbara Weinstein, and Nimrod Shiman. 2020. 'How Do Medical Masks Degrade Speech Reception?' *Hearing Review* 27(5):8–9.

Gomez, Pablo, Roger Ratcliff, and Manuel Perea. 2007. 'A Model of the Go/No-Go Task.' *Journal of Experimental Psychology: General* 136(3):389–413. doi: 10.1037/0096-3445.136.3.389.

Gosselin, Penny Anderson, and Jean-Pierre Gagné. 2010. 'Use of a Dual-Task Paradigm to Measure Listening Effort'. *Canadian Journal of Speech-Language Pathology and Audiology* 34(1):43–51.

Granholm, E., R. F. Asarnow, A. J. Sarkin, and K. L. Dykes. 1996. 'Pupillary Responses Index Cognitive Resource Limitations.' *Psychophysiology* 33(4):457–61.

Grömping, Ulrike. 2006. 'Relative Importance for Linear Regression in *R* : The Package Relaimpo.' *Journal of Statistical Software* 17(1):1–27. doi: https://doi.org/10.18637/jss.v017.i01.

Grote, Helen, and Fizz Izagaren. 2020. ‘Covid-19: The Communication Needs of D/Deaf Healthcare Workers and Patients Are Being Forgotten’. *BMJ* 369:m2372. doi: 10.1136/bmj.m2372.

Gulliver, S. R., and G. Ghinea. 2003. ‘How Level and Type of Deafness Affects User Perception of Multimedia Video Clips’. *Universal Access in the Information Society* 2(4):374–86. doi: https://doi.org/10.1007/s10209-003- 0067-5.

Gunawan, David, Guy E. Hawkins, Minh-Ngoc Tran, Robert Kohn, and Scott Brown. 2020. ‘New Estimation Approaches for the Hierarchical Linear Ballistic Accumulator Model’. *ArXiv:1806.10089 [Stat]*.

Hällgren, Mathias, Birgitta Larsby, Björn Lyxell, and Stig Arlinger. 2001. ‘Evaluation of a Cognitive Test Battery in Young and Elderly Normal-Hearing and Hearing-Impaired Persons’. *Journal of the American Academy of Audiology* 12(07):357–70. doi: 10.1055/s-0042-1745620.

Harrison, Samantha, and Douglas Hartley. 2019. ‘Shedding Light On The Human Auditory Cortex: A Review Of The Advances In Near Infrared Spectroscopy (NIRS)’. *Reports in Medical Imaging* Volume 12:31–42. doi: 10.2147/RMI.S174633.

Hart, Sandra G., and Lowell E. Staveland. 1988. ‘Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research’. Pp. 139–83 in *Advances in Psychology*. Vol. 52.

Hast, Anne, Luisa Schlücker, Frank Digeser, Tim Liebscher, and Ulrich Hoppe. 2015. ‘Speech Perception of Elderly Cochlear Implant Users Under Different Noise Conditions’. *Otology & Neurotology* 36(10):1638–43. doi: 10.1097/MAO.0000000000000883.

Haumann, Sabine, Volker Hohmann, Markus Meis, Tobias Herzke, Thomas Lenarz, and Andreas Büchner. 2012. ‘Indication Criteria for Cochlear Implants and Hearing Aids: Impact of Audiological and Non-Audiological Findings’. *Audiology Research* 2(1). doi: 10.4081/audiores.2012.e12.

Hazrati, Oldooz, and Philipos C. Loizou. 2012. ‘The Combined Effects of Reverberation and Noise on Speech Intelligibility by Cochlear Implant Listeners’. *International Journal of Audiology* 51(6):437–43. doi: 10.3109/14992027.2012.658972.

Heathcote, Andrew, and Jonathon Love. 2012. ‘Linear Deterministic Accumulator Models of Simple Choice’. *Frontiers in Psychology* 3. doi: 10.3389/fpsyg.2012.00292.

Heitz, Richard P. 2014. ‘The Speed-Accuracy Tradeoff: History, Physiology, Methodology, and Behavior’. *Frontiers in Neuroscience* 8. doi: 10.3389/fnins.2014.00150.

Herold, Fabian, Patrick Wiegel, Felix Scholkmann, and Notger Müller. 2018. ‘Applications of Functional Near-Infrared Spectroscopy (FNIRS) Neuroimaging in Exercise–Cognition Science: A Systematic, Methodology-Focused Review’. *Journal of Clinical Medicine* 7(12):466. doi: 10.3390/jcm7120466.

Herrmann, Björn, and Ingrid S. Johnsrude. 2020. ‘A Model of Listening Engagement (MoLE)’. *Hearing Research*. doi: https://doi.org/10.1016/j.heares.2020.108016.

Hetu, R., L. Riverin, N. Lalande, L. Getty, and C. St-Cyr. 1988. ‘Qualitative Analysis of the Handicap Associated with Occupational Hearing Loss.’ *British Journal of Audiology* 22(4):251–64.

Hockey, Robert. 2013. *The Psychology of Fatigue: Work, Effort and Control*. Cambridge: Cambridge University Press.

van der Hoek-Snieders, Hanneke E. M., Monique Boymans, Bas Sorgdrager, and Wouter A. Dreschler. 2020. ‘Factors Influencing the Need for Recovery in Employees with Hearing Loss: A Cross-Sectional Study of Health Administrative Data’. *International Archives of Occupational and Environmental Health*. doi: 10.1007/s00420-020-01556-z.

Holm, Sture. 1979. ‘A Simple Sequentially Rejective Multiple Test Procedure’. *Scandinavian Journal of Statistics* 6(2):65–70.

Hopstaken, Jesper F., Dimitri van der Linden, Arnold B. Bakker, and Michiel A. J. Kompier. 2015. ‘A Multifaceted Investigation of the Link between Mental Fatigue and Task Disengagement: Mental Fatigue and Task Disengagement’. *Psychophysiology* 52(3):305–15. doi: 10.1111/psyp.12339.

Hopstaken, Jesper F., Dimitri van der Linden, Arnold B. Bakker, and Michiel A.J. Kompier. 2015. ‘The Window of My Eyes: Task Disengagement and Mental Fatigue Covary with Pupil Dynamics’. *Biological Psychology* 110:100–106. doi: 10.1016/j.biopsycho.2015.06.013.

Hornsby, Benjamin W. Y. 2013. ‘The Effects of Hearing Aid Use on Listening Effort and Mental Fatigue Associated With Sustained Speech Processing Demands’. *Ear & Hearing* 34(5):523–34. doi: 10.1097/AUD.0b013e31828003d8.

Hornsby, Benjamin W. Y., Graham Naylor, and Fred H. Bess. 2016. ‘A Taxonomy of Fatigue Concepts and Their Relation to Hearing Loss.’ *Ear and Hearing* 37:136S-144S. doi: 10.1097/AUD.0000000000000289.

Houben, Rolph, Maaike van Doorn-Bierman, and Wouter A. Dreschler. 2013. ‘Using Response Time to Speech as a Measure for Listening Effort’. *International Journal of Audiology* 52(11):753–61. doi: 10.3109/14992027.2013.832415.

Hughes, Meredith M., Jared A. Linck, Anita R. Bowles, Joel T. Koeth, and Michael F. Bunting. 2014. ‘Alternatives to Switch-Cost Scoring in the Task-Switching Paradigm: Their Reliability and Increased Validity’. *Behavior Research Methods* 46(3):702–21. doi: 10.3758/s13428-013-0411-5.

Hughes, Sarah E., Hayley A. Hutchings, Frances L. Rapport, Catherine M. McMahon, and Isabelle Boisvert. 2018. ‘Social Connectedness and Perceived Listening Effort in Adult Cochlear Implant Users: A Grounded Theory to Establish Content Validity for a New Patient-Reported Outcome Measure’. 13.

Huppert, Theodore J., Solomon G. Diamond, Maria A. Franceschini, and David A. Boas. 2009. ‘HomER: A Review of Time-Series Analysis Methods for near-Infrared Spectroscopy of the Brain’. *Applied Optics* 48(10):D280. doi: 10.1364/AO.48.00D280.

Hussein, Asmaa Bakr, Reham Mamdouh Lasheen, Afaf Ahmed Emara, and Trandil El Mahallawi. 2022. ‘Listening Effort in Patients with Sensorineural Hearing Loss with and without Hearing Aids’. *The Egyptian Journal of Otolaryngology* 38(1):99. doi: 10.1186/s43163-022-00288-w.

Hygge, Staffan, Jerker Rönnberg, Birgitta Larsby, and Stig Arlinger. 1992. ‘Normal-Hearing and Hearing- Impaired Subjects’ Ability to Just Follow Conversation in Competing Speech, Reversed Speech, and Noise Backgrounds’. *Journal of Speech, Language, and Hearing Research* 35(1):208–15. doi: 10.1044/jshr.3501.208.

Hyland, P., M. Shevlin, O. McBride, J. Murphy, T. Karatzias, R. P. Bentall, A. Martinez, and F. Vallières. 2020. ‘Anxiety and Depression in the Republic of Ireland during the COVID-19 Pandemic’. *Acta Psychiatrica Scandinavica* 142(3):249–56. doi: 10.1111/acps.13219.

Hyndman, Rob J. 1996. ‘Computing and Graphing Highest Density Regions’. *The American Statistician* 50(2):120–26. doi: https://doi.org/10.2307/2684423.

Hyseni Duraku, Zamira, and Linda Hoxha. 2020. *Impact of the COVID-19 Pandemic on Education and Wellbeing*.

Ideas For Ears. 2018. ‘Phone Calls Are Causing Frustration and Stress’. Retrieved 23 February 2021 (https://www.ideasforears.org.uk/blog/phone-calls-are-causing-frustration-and-stress/).

Ideas for Ears Ltd. 2020. *Coronavirus & the Unequal Impact on People with Hearing Loss*.

Jasper, H. H. 1958. ‘The Ten-Twenty Electrode System of the International Federation’. *Electroencephalography and Clinical Neurophysiology* 10:371–75.

Jelinek Lewis, Margaret S., and Dorothy W. Jackson. 2001. ‘Television Literacy: Comprehension of Program Content Using Closed Captions for the Deaf’. *Journal of Deaf Studies and Deaf Education* 6(1):43–53. doi: 10.1093/deafed/6.1.43.

Jensema, Carl J., Sameh El Sharkawy, Ramalinga Sarma Danturthi, Robert Burch, and David Hsu. 2000. ‘Eye Movement Patterns of Captioned Television Viewers’. *American Annals of the Deaf* 145(3):275–85. doi: 10.1353/aad.2012.0093.

Jiang, Manyu. 2020. ‘The Reason Zoom Calls Drain Your Energy’. *BBC Worklife*, April 22.

Jiwani, Salima, Blake C. Papsin, and Karen A. Gordon. 2016. ‘Early Unilateral Cochlear Implantation Promotes Mature Cortical Asymmetries in Adolescents Who Are Deaf’. *Human Brain Mapping* 37(1):135–52. doi: https://doi.org/10.1002/hbm.23019.

Jobsis, F. 1977. ‘Noninvasive, Infrared Monitoring of Cerebral and Myocardial Oxygen Sufficiency and Circulatory Parameters’. *Science* 198(4323):1264–67. doi: 10.1126/science.929199.

Johnson, Jani, Jingjing Xu, Robyn Cox, and Paul Pendergraft. 2015. ‘A Comparison of Two Methods for Measuring Listening Effort As Part of an Audiologic Test Battery’. *American Journal of Audiology* 24(3):419–31. doi: 10.1044/2015_AJA-14-0058.

Jones, Nicholas R., Zeshan U. Qureshi, Robert J. Temple, Jessica P. J. Larwood, Trisha Greenhalgh, and Lydia Bourouiba. 2020. ‘Two Metres or One: What Is the Evidence for Physical Distancing in Covid-19?’ *BMJ* m3223. doi: 10.1136/bmj.m3223.

Joshi, Siddhartha, Yin Li, Rishi M. Kalwani, and Joshua I. Gold. 2016. ‘Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex’. *Neuron* 89(1):221–34. doi: 10.1016/j.neuron.2015.11.028.

Kahneman, Daniel. 1973. *Attention and Effort*. Englewood Cliffs, N.J: Prentice-Hall,Inc.

Kane, Gerald C., Rich Nanda, Anh Phillips, and Jonathan Copulsky. 2021. ‘Redesigning the Post-Pandemic Workplace’. *MIT SLOAN MANAGEMENT REVIEW*, February 10, 4.

Kaplan Neeman, Ricky, Ilan Roziner, and Chava Muchnik. 2022. ‘A Clinical Paradigm for Listening Effort Assessment in Middle-Aged Listeners’. *Frontiers in Psychology* 13:820227. doi: 10.3389/fpsyg.2022.820227.

Karatekin, Canan. 2004. ‘Development of Attentional Allocation in the Dual Task Paradigm.’ *International Journal of Psychophysiology* 52(1):7–21. doi: 10.1016/j.ijpsycho.2003.12.002.

Katahira, Kentaro. 2016. ‘How Hierarchical Models Improve Point Estimates of Model Parameters at the Individual Level’. *Journal of Mathematical Psychology* 73:37–58. doi: 10.1016/j.jmp.2016.03.007.

Kitterick, Pádraig T., and Laura Lucas. 2016. ‘Predicting Speech Perception Outcomes Following Cochlear Implantation in Adults with Unilateral Deafness or Highly Asymmetric Hearing Loss’. *Cochlear Implants International* 17(sup1):51–54. doi: 10.1080/14670100.2016.1155806.

Kniffin, Kevin M., Jayanth Narayanan, Frederik Anseel, John Antonakis, Susan P. Ashford, Arnold B. Bakker, Peter Bamberger, Hari Bapuji, Devasheesh P. Bhave, Virginia K. Choi, Stephanie J. Creary, Evangelia Demerouti, Francis J. Flynn, Michele J. Gelfand, Lindred L. Greer, Gary Johns, Selin Kesebir, Peter G. Klein, Sun Young Lee, Hakan Ozcelik, Jennifer Louise Petriglieri, Nancy P. Rothbard, Cort W. Rudolph, Jason D. Shaw, Nina Sirola, Connie R. Wanberg, Ashley Whillans, Michael P. Wilmot, and Mark van Vugt. 2021. ‘COVID-19 and the Workplace: Implications, Issues, and Insights for Future Research and Action.’ *American Psychologist* 76(1):63–77. doi: 10.1037/amp0000716.

Koelewijn, Thomas, José A. P. van Haastrecht, and Sophia E. Kramer. 2018. ‘Pupil Responses of Adults With Traumatic Brain Injury During Processing of Speech in Noise’. *Trends in Hearing* 22:233121651881144. doi: 10.1177/2331216518811444.

Koelewijn, Thomas, Hilde de Kluiver, Barbara G Shinn-Cunningham, Adriana A. Zekveld, and Sophia E. Kramer. 2015. ‘The Pupil Response Reveals Increased Listening Effort When It Is Difficult to Focus Attention’. *Hearing Research* 323:81–90. doi: 10.1016/j.heares.2015.02.004.

Koelewijn, Thomas, Hilde de Kluiver, Barbara G. Shinn-Cunningham, Adriana A. Zekveld, and Sophia E. Kramer. 2015. ‘The Pupil Response Reveals Increased Listening Effort When It Is Difficult to Focus Attention’. *Hearing Research* 323:81–90. doi: 10.1016/j.heares.2015.02.004.

Koelewijn, Thomas, Niek J. Versfeld, and Sophia E. Kramer. 2017. ‘Effects of Attention on the Speech Reception Threshold and Pupil Response of People with Impaired and Normal Hearing’. *Hearing Research* 354:56–63. doi: 10.1016/j.heares.2017.08.006.

Kouneiher, Frédérique, Sylvain Charron, and Etienne Koechlin. 2009. ‘Motivation and Cognitive Control in the Human Prefrontal Cortex’. *Nature Neuroscience* 12(7):939–45. doi: 10.1038/nn.2321.

Kramer, Sophia E., Theo S. Kapteyn, and Tammo Houtgast. 2006. ‘Occupational Performance: Comparing Normally-Hearing and Hearing-Impaired Employees Using the Amsterdam Checklist for Hearing and Work.’ *International Journal of Audiology* 45(9):503–12. doi: 10.1080/14992020600754583.

Kramer, Sophia E., Charlotte E. Teunissen, and Adriana A. Zekveld. 2016. ‘Cortisol, Chromogranin A, and Pupillary Responses Evoked by Speech Recognition Tasks in Normally Hearing and Hard-of-Hearing Listeners: A Pilot Study.’ *Ear and Hearing* 37 Suppl 1:126S-35S. doi: 10.1097/AUD.0000000000000311.

Kramer, Sophia E., Adriana A. Zekveld, and Tammo Houtgast. 2009. ‘Measuring Cognitive Factors in Speech Comprehension: The Value of Using the Text Reception Threshold Test as a Visual Equivalent of the SRT Test’. *Scandinavian Journal of Psychology* 50(5):507–15. doi: 10.1111/j.1467-9450.2009.00747.x.

Kressner, Abigail Anne, Adam Westermann, and Jörg M. Buchholz. 2018. ‘The Impact of Reverberation on Speech Intelligibility in Cochlear Implant Recipients’. *The Journal of the Acoustical Society of America* 144(2):1113–22. doi: 10.1121/1.5051640.

Krueger, Melanie, Michael Schulte, Thomas Brand, and Inga Holube. 2017. ‘Development of an Adaptive Scaling Method for Subjective Listening Effort’. *The Journal of the Acoustical Society of America* 141(6):4680–93. doi: 10.1121/1.4986938.

Kubinec, Robert. 2022. ‘Ordered Beta Regression: A Parsimonious, Well-Fitting Model for Continuous Data with Lower and Upper Bounds’. *Political Analysis*. doi: 10.31235/osf.io/2sx6y.

Laeng, Bruno, Sylvain Sirois, and Gustaf Gredebäck. 2012. ‘Pupillometry: A Window to the Preconscious?’ *Perspectives on Psychological Science* 7(1):18–27. doi: 10.1177/1745691611427305.

Larsby, Birgitta, Mathias Hallgren, Bjorn Lyxell, and Stig Arlinger. 2005. ‘Cognitive Performance and Perceived Effort in Speech Processing Tasks: Effects of Different Noise Backgrounds in Normal-Hearing and Hearing-Impaired Subjects.’ *International Journal of Audiology* 44(3):131–43.

Lau, Marcy K., Candace Hicks, Tobias Kroll, and Steven Zupancic. 2019. ‘Effect of Auditory Task Type on Physiological and Subjective Measures of Listening Effort in Individuals With Normal Hearing.’ *Journal of Speech, Language, and Hearing Research : JSLHR* 62(5):1549–60. doi: 10.1044/2018_JSLHR-H-17-0473.

Lavie, Nilli. 1995. ‘Perceptual Load as a Necessary Condition for Selective Attention’. *Journal of Experimental Psychology Human Perception and Performance* 21(3):451–68. doi: doi: 10.1037//0096-1523.21.3.451.

Lawrence, Rachael J., Ian M. Wiggins, Carly A. Anderson, Jodie Davies-Thompson, and Douglas E. H. Hartley. 2018. ‘Cortical Correlates of Speech Intelligibility Measured Using Functional Near-Infrared Spectroscopy (FNIRS)’. *Hearing Research* 370:53–64. doi: 10.1016/j.heares.2018.09.005.

Lawrence, Rachael J., Ian M. Wiggins, Jessica C. Hodgson, and Douglas E. H. Hartley. 2021. ‘Evaluating Cortical Responses to Speech in Children: A Functional near-Infrared Spectroscopy (FNIRS) Study’. *Hearing Research* 401:108155. doi: 10.1016/j.heares.2020.108155.

Lee, Michael D., and Eric-Jan Wagenmakers. 2014. *Bayesian Cognitive Modeling: A Practical Course*. Cambridge University Press.

Li, Xiangrui, Zhen Liang, Mario Kleiner, and Zhong-Lin Lu. 2010. ‘RTbox: A Device for Highly Accurate Response Time Measurements’. *Behavior Research Methods* 42(1):212–25. doi: 10.3758/BRM.42.1.212.

Liesefeld, Heinrich René, and Markus Janczyk. 2019. ‘Combining Speed and Accuracy to Control for Speed-Accuracy Trade-Offs(?)’. *Behavior Research Methods* 51(1):40–60. doi: https://doi.org/10.3758/s13428-018-1076-x.

Lin, Frank R., E. Jeffrey Metter, Richard J. O’Brien, Susan M. Resnick, Alan B. Zonderman, and Luigi Ferrucci. 2011. ‘Hearing Loss and Incident Dementia’. *Archives of Neurology* 68(2). doi: 10.1001/archneurol.2010.362.

Lin, Frank R., Kristine Yaffe, Jin Xia, Qian-Li Xue, Tamara B. Harris, Elizabeth Purchase-Helzner, Suzanne Satterfield, Hilsa N. Ayonayon, Luigi Ferrucci, Eleanor M. Simonsick, and for the Health ABC Study Group. 2013. ‘Hearing Loss and Cognitive Decline in Older Adults’. *JAMA Internal Medicine* 173(4):293. doi: 10.1001/jamainternmed.2013.1868.

Lindsley, William G., John D. Noti, Francoise M. Blachere, Jonathan V. Szalajda, and Donald H. Beezhold. 2014. ‘Efficacy of Face Shields Against Cough Aerosol Droplets from a Cough Simulator’. *Journal of Occupational and Environmental Hygiene* 11(8):509–18. doi: 10.1080/15459624.2013.877591.

Liu, Xiao-Wei, Da-Gang Lu, and Pierre C. J. Hoogenboom. 2017. ‘Hierarchical Bayesian Fatigue Data Analysis’. *International Journal of Fatigue* 100:418–28. doi: 10.1016/j.ijfatigue.2017.03.043.

Livingston, Gill, Jonathan Huntley, Andrew Sommerlad, David Ames, Clive Ballard, Sube Banerjee, Carol Brayne, Alistair Burns, Jiska Cohen-Mansfield, Claudia Cooper, Sergi G. Costafreda, Amit Dias, Nick Fox, Laura N. Gitlin, Robert Howard, Helen C. Kales, Mika Kivimäki, Eric B. Larson, Adesola Ogunniyi, Vasiliki Orgeta, Karen Ritchie, Kenneth Rockwood, Elizabeth L. Sampson, Quincy Samus, Lon S. Schneider, Geir Selbæk, Linda Teri, and Naaheed Mukadam. 2020. ‘Dementia Prevention, Intervention, and Care: 2020 Report of the Lancet Commission’. *The Lancet* S0140673620303676. doi: 10.1016/S0140-6736(20)30367-6.

Lorenzo-Seva, Urbano, and Jos M. F. ten Berge. 2006. ‘Tucker’s Congruence Coefficient as a Meaningful Index of Factor Similarity’. *Methodology* 2(2):57–64. doi: 10.1027/1614-2241.2.2.57.

Lunner, Thomas. 2003. 'Cognitive Function in Relation to Hearing Aid Use.' *International Journal of Audiology* 42 Suppl 1:S49-58. doi: 10.3109/14992020309074624.

Lunner, Thomas, Mary Rudner, Tove Rosenbom, Jessica Ågren, and Elaine Hoi Ning Ng. 2016. 'Using Speech Recall in Hearing Aid Fitting and Outcome Evaluation Under Ecological Test Conditions': *Ear and Hearing* 37:145S-154S. doi: 10.1097/AUD.0000000000000294.

Ly, Alexander, Udo Boehm, Andrew Heathcote, Brandon M. Turner, Birte Forstmann, Maarten Marsman, and Dora Matzke. 2017. 'A Flexible and Efficient Hierarchical Bayesian Approach to the Exploration of Individual Differences in Cognitive-Model-Based Neuroscience'. Pp. 467–79 in *Computational Models of Brain and Behavior*, edited by A. A. Moustafa. Chichester, UK: John Wiley & Sons, Ltd.

Ly, Alexander, Maarten Marsman, and Eric-Jan Wagenmakers. 2018. 'Analytic Posteriors for Pearson's Correlation Coefficient'. *Statistica Neerlandica* 72(1):4–13. doi: doi:10.1111/stan.12111.

Mackersie, Carol L., and Natalie Calderon-Moultrie. 2016. 'Autonomic Nervous System Reactivity During Speech Repetition Tasks: Heart Rate Variability and Skin Conductance.' *Ear and Hearing* 37 Suppl 1:118S-25S. doi: 10.1097/AUD.0000000000000305.

Magee, Michelle, Courtney Lewis, Gustavo Noffs, Hannah Reece, Jess C. S. Chan, Charissa J. Zaga, Camille Paynter, Olga Birchall, Sandra Rojas Azocar, Angela Ediriweera, Katherine Kenyon, Marja W. Caverlé, Benjamin G. Schultz, and Adam P. Vogel. 2020. 'Effects of Face Masks on Acoustic Analysis and Speech Perception: Implications for Peri-Pandemic Protocols'. *The Journal of the Acoustical Society of America* 148(6):3562–68. doi: 10.1121/10.0002873.

Maru, Devina, Jack Stancel-Lewis, Graham Easton, and William EJ Leverton. 2021. 'Communicating with People with Hearing Loss: COVID-19 and Beyond'. *BJGP Open* BJGPO.2020.0174. doi: 10.3399/BJGPO.2020.0174.

Masalimova, Alfiya R., Elena L. Ryazanova, Larisa I. Tararina, Ekaterina G. Sokolova, Yuliya B. Ikrennikova, Svetlana V. Efimushkina, and Tatiana I. Shulga. 2021. 'Distance Learning Hybrid Format for University Students in Post-Pandemic Perspective: Collaborative Technologies Aspect'. *Cypriot Journal of Educational Sciences* 16(1):389–95. doi: 10.18844/cjes.v16i1.5536.

Matthen, Mohan. 2016. 'Effort and Displeasure in People Who Are Hard of Hearing': *Ear and Hearing* 37:28S-34S. doi: 10.1097/AUD.0000000000000292.

Mattys, Sven L., Matthew H. Davis, Ann R. Bradlow, and Sophie K. Scott. 2012. 'Speech Recognition in Adverse Conditions: A Review'. *Language and Cognitive Processes* 27(7–8):953–78. doi: 10.1080/01690965.2012.705006.

Mattys, Sven L., and Lukas Wiget. 2011. 'Effects of Cognitive Load on Speech Recognition'. *Journal of Memory and Language* 65(2):145–60. doi: 10.1016/j.jml.2011.04.004.

Mayhew, John, David Johnston, Jason Berwick, Myles Jones, Peter Coffey, and Ying Zheng. 2000. 'Spectroscopic Analysis of Neural Activity in Brain: Increased Oxygen

Consumption Following Activation of Barrel Cortex'. *NeuroImage* 12(6):664–75. doi: 10.1006/nimg.2000.0656.

McCoy, Sandra L., Patricia A. Tun, L. Clarke Cox, Marianne Colangelo, Raj A. Stewart, and Arthur Wingfield. 2005. 'Hearing Loss and Perceptual Effort: Downstream Effects on Older Adults' Memory for Speech.' *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology* 58(1):22–33. doi: 10.1080/02724980443000151.

McGarrigle, Ronan, Sarah Knight, Lyndon Rakusen, Jason Geller, and Sven Mattys. 2021. 'Older Adults Show a More Sustained Pattern of Effortful Listening Than Young Adults'. *Psychology and Aging* 36(4):504–19. doi: https://doi.org/10.1037/pag0000587.

McGarrigle, Ronan, Kevin J. Munro, Piers Dawes, Andrew J. Stewart, David R. Moore, Johanna G. Barry, and Sygal Amitay. 2014a. 'Listening Effort and Fatigue: What Exactly Are We Measuring? A British Society of Audiology Cognition in Hearing Special Interest Group "White Paper"'. *International Journal of Audiology* 53(7):433–45. doi: 10.3109/14992027.2014.890296.

McGarrigle, Ronan, Kevin J. Munro, Piers Dawes, Andrew J. Stewart, David R. Moore, Johanna G. Barry, and Sygal Amitay. 2014b. 'Listening Effort and Fatigue: What Exactly Are We Measuring? A British Society of Audiology Cognition in Hearing Special Interest Group "White Paper"'. *International Journal of Audiology* 53(7):433–45. doi: 10.3109/14992027.2014.890296.

McGarrigle, Ronan, Lyndon Rakusen, and Sven Mattys. 2021. 'Effortful Listening under the Microscope: Examining Relations between Pupillometric and Subjective Markers of Effort and Tiredness from Listening'. *Psychophysiology* 58(1). doi: 10.1111/psyp.13703.

McGettigan, Carolyn, Samuel Evans, Stuart Rosen, Zarinah K. Agnew, Poonam Shah, and Sophie K. Scott. 2012. 'An Application of Univariate and Multivariate Approaches in FMRI to Quantifying the Hemispheric Lateralization of Acoustic and Linguistic Processes'. *Journal of Cognitive Neuroscience* 24(3):636–52. doi: 10.1162/jocn_a_00161.

McKay, Colette M., Adnan Shah, Abd-Krim Seghouane, Zhou Xin, William Cross, and Ruth Litovsky. 2016. 'Connectivity in Language Areas of the Brain in Cochlear Implant Users as Revealed by FNIRS'. *Advances in Experimental Medicine and Biology* 894:327–35. doi: 10.1007/978-3-319-25474-6_34.

McRackan, Theodore R., Brittany N. Hand, Cochlear Implant Quality of Life Development Consortium, Craig A. Velozo, and Judy R. Dubno. 2019. 'Cochlear Implant Quality of Life (CIQOL): Development of a Profile Instrument (CIQOL-35 Profile) and a Global Measure (CIQOL-10 Global)'. *Journal of Speech, Language, and Hearing Research* 62(9):3554–63. doi: 10.1044/2019_JSLHR-H-19-0142.

McRackan, Theodore R., Brittany N. Hand, Craig A. Velozo, and Judy R. Dubno. 2019. 'Association of Demographic and Hearing-Related Factors With Cochlear Implant–Related Quality of Life'. *JAMA Otolaryngology–Head & Neck Surgery* 145(5):422. doi: 10.1001/jamaoto.2019.0055.

Michielsen, Helen J., Jolanda De Vries, Guus L. Van Heck, Fons J. R. Van de Vijver, and Klaas Sijtsma. 2004. ‘Examination of the Dimensionality of Fatigue: The Construction of the Fatigue Assessment Scale (FAS)’. *European Journal of Psychological Assessment* 20(1):39–48. doi: 10.1027/1015-5759.20.1.39.

Mick, Paul, Ichiro Kawachi, and Frank R. Lin. 2014. ‘The Association between Hearing Loss and Social Isolation in Older Adults’. *Otolaryngology–Head and Neck Surgery* 150(3):378–84. doi: 10.1177/0194599813518021.

Moberly, Aaron C., Kara J. Vasil, and Christin Ray. 2020. ‘Visual Reliance during Speech Recognition in Cochlear Implant Users and Candidates’. *Journal of the American Academy of Audiology* 31(1):30–39. doi: 10.3766/jaaa.18049.

Moore, Brian C. J. 2008. ‘The Role of Temporal Fine Structure Processing in Pitch Perception, Masking, and Speech Perception for Normal-Hearing and Hearing-Impaired People’. *Journal of the Association for Research in Otolaryngology* 9(4):399–406. doi: 10.1007/s10162-008-0143-x.

Moore, Travis M., and Erin M. Picou. 2018. ‘A Potential Bias in Subjective Ratings of Mental Effort’. *Journal of Speech, Language, and Hearing Research* 61(9):2405–21. doi: 10.1044/2018_JSLHR-H-17-0451.

Mushtaq, Faizah, Ian M. Wiggins, Padraig T. Kitterick, Carly A. Anderson, and Douglas E. H. Hartley. 2019. ‘Evaluating Time-Reversed Speech and Signal-Correlated Noise as Auditory Baselines for Isolating Speech-Specific Processing Using FNIRS’. *PLoS ONE* 14(7):e0219927. doi: https://doi.org/10.1371/journal. pone.0219927.

Mushtaq, Faizah, Ian M. Wiggins, Pádraig T. Kitterick, Carly A. Anderson, and Douglas E. H. Hartley. 2021. ‘Investigating Cortical Responses to Noise-Vocoded Speech in Children with Normal Hearing Using Functional Near-Infrared Spectroscopy (FNIRS)’. *Journal of the Association for Research in Otolaryngology* 22(6):703–17. doi: 10.1007/s10162-021-00817-z.

Nachtegaal, Janneke, Dirk J. Kuik, Johannes R. Anema, S. Theo Goverts, Joost M. Festen, and Sophia E. Kramer. 2009. ‘Hearing Status, Need for Recovery after Work, and Psychosocial Work Characteristics: Results from an Internet-Based National Survey on Hearing’. *International Journal of Audiology* 48(10):684–91. doi: 10.1080/14992020902962421.

Naito, Yasushi, Ichiro Tateya, Nobuya Fujiki, Shigeru Hirano, Koichi Ishizu, Yasuhiro Nagahama, Hidenao Fukuyama, and Hisayoshi Kojima. 2000. ‘Increased Cortical Activation during Hearing of Speech in Cochlear Implant Users’. *Hearing Research* 143(1–2):139–46. doi: 10.1016/S0378-5955(00)00035-6.

Naseer, Noman, and Keum-Shik Hong. 2015. ‘FNIRS-Based Brain-Computer Interfaces: A Review’. *Frontiers in Human Neuroscience* 9. doi: 10.3389/fnhum.2015.00003.

Naylor, Graham, Louise A. Burke, and Jack A. Holman. 2020. ‘Covid-19 Lockdown Affects Hearing Disability and Handicap in Diverse Ways: A Rapid Online Survey Study’. *Ear & Hearing* 41(6):1442–49. doi: 10.1097/AUD.0000000000000948.

Naylor, Graham, Thomas Koelewijn, Adriana A. Zekveld, and Sophia E. Kramer. 2018. 'The Application of Pupillometry in Hearing Science to Assess Listening Effort'. *Trends in Hearing* 22:233121651879943. doi: 10.1177/2331216518799437.

NDC. 2020. 'How to Host Effective and Accessible Online Meetings with Deaf Participants'. *National Deaf Center*. Retrieved 25 January 2022 (https://www.nationaldeafcenter.org/news/how-host-effective-and-accessible-online-meetings-deaf-participants).

Needleman, Alyssa R., and Carl C. Crandellt. 1995. 'Speech Recognition in Noise by Hearing-Impaired and Noise-Masked Normal-Hearing Listeners'. *Journal of the American Academy of Audiology* 6(6):414–24.

Newman, Craig W., Barbara E. Weinstein, Gary P. Jacobson, and Gerald A. Hug. 1991. 'Test-Retest Reliability of the Hearing Handicap Inventory for Adults': *Ear and Hearing* 12(5):355–57. doi: 10.1097/00003446-199110000-00009.

NICE. 2019. *Cochlear Implants for Children and Adults with Severe to Profound Deafness*. Ref: TA566. National Institute for Health and Care Excellence (NICE). Technology appraisal guidance 566.

Nilsson, Håkan, Jörg Rieskamp, and Eric-Jan Wagenmakers. 2011. 'Hierarchical Bayesian Parameter Estimation for Cumulative Prospect Theory'. *Journal of Mathematical Psychology* 55:84–93.

Noah, J. Adam, Yumie Ono, Yasunori Nomoto, Sotaro Shimada, Atsumichi Tachibana, Xian Zhang, Shaw Bronner, and Joy Hirsch. 2015. 'FMRI Validation of FNIRS Measurements During a Naturalistic Task'. *Journal of Visualized Experiments : JoVE* (100):e52116–e52116. doi: 10.3791/52116.

Noble, William, Niels Søgaard Jensen, Graham Naylor, Navjot Bhullar, and Michael A. Akeroyd. 2013. 'A Short Form of the Speech, Spatial and Qualities of Hearing Scale Suitable for Clinical Use: The SSQ12'. *International Journal of Audiology* 52(6):409–12. doi: 10.3109/14992027.2013.781278.

Ohlenforst, Barbara, Adriana A. Zekveld, Elise P. Jansma, Yang Wang, Graham Naylor, Artur Lorens, Thomas Lunner, and Sophia E. Kramer. 2017. 'Effects of Hearing Impairment and Hearing Aid Amplification on Listening Effort: A Systematic Review'. *Ear and Hearing* 38(3):267–81. doi: 10.1097/AUD.0000000000000396.

Oldfield, R. C. 1971. 'The Assessment and Analysis of Handedness: The Edinburgh Inventory.' *Neuropsychologia* 9(1):97–113. doi: 10.1016/0028-3932(71)90067-4.

Olds, Cristen, Luca Pollonini, Homer Abaya, Jannine Larky, Megan Loy, Heather Bortfeld, Michael S. Beauchamp, and John S. Oghalai. 2016. 'Cortical Activation Patterns Correlate with Speech Understanding After Cochlear Implantation'. *Ear & Hearing* 37(3):e160–72. doi: 10.1097/AUD.0000000000000258.

O'Neill, Erin R., Morgan N. Parke, Heather A. Kreft, and Andrew J. Oxenham. 2021. 'Role of Semantic Context and Talker Variability in Speech Perception of Cochlear-Implant Users and Normal-Hearing Listeners'. *The Journal of the Acoustical Society of America* 149(2):1224–39. doi: 10.1121/10.0003532.

Paas, Fred, Alexander Renkl, and John Sweller. 2003. ‘Cognitive Load Theory and Instructional Design: Recent Developments’. *Educational Psychologist* 38(1):1–4. doi: 10.1207/S15326985EP3801_1.

Pals, Carina, Anastasios Sarampalis, and Deniz Başkent. 2013. ‘Listening Effort With Cochlear Implant Simulations’. *Journal of Speech, Language, and Hearing Research* 56(4):1075–84. doi: 10.1044/1092-4388(2012/12-0074).

Pals, Carina, Anastasios Sarampalis, Andy Beynon, Thomas Stainsby, and Deniz Başkent. 2020. ‘Effect of Spectral Channels on Speech Recognition, Comprehension, and Listening Effort in Cochlear-Implant Users’. *Trends in Hearing* 24:1–15. doi: 10.1177/2331216520904617.

Pang, Jermy, Elizabeth Francis Beach, Megan Gilliver, and Ingrid Yeend. 2019. ‘Adults Who Report Difficulty Hearing Speech in Noise: An Exploration of Experiences, Impacts and Coping Strategies.’ *International Journal of Audiology* 1–10. doi: 10.1080/14992027.2019.1670363.

Pantev, C., A. Dinnesen, B. Ross, A. Wollbrink, and A. Knief. 2006. ‘Dynamics of Auditory Plasticity after Cochlear Implantation: A Longitudinal Study.’ *Cerebral Cortex (New York, N.Y. : 1991)* 16(1):31–36. doi: 10.1093/cercor/bhi081.

Patil, Amol V., Javad Safaie, Hamid Abrishami Moghaddam, Fabrice Wallois, and Reinhard Grebe. 2011. ‘Experimental Investigation of NIRS Spatial Sensitivity’. *Biomedical Optics Express* 2(6):1478. doi: 10.1364/BOE.2.001478.

Paulus, Martin P., Corianne Rogalsky, Alan Simmons, Justin S. Feinstein, and Murray B. Stein. 2003. ‘Increased Activation in the Right Insula during Risk-Taking Decision Making Is Related to Harm Avoidance and Neuroticism’. *NeuroImage* 19(4):1439–48. doi: 10.1016/S1053-8119(03)00251-9.

Peelle, Jonathan E. 2018. ‘Listening Effort: How the Cognitive Consequences of Acoustic Challenge Are Reflected in Brain and Behavior’. *Ear and Hearing* 39(2):204–14. doi: 10.1097/AUD.0000000000000494.

Peelle, Jonathan E., Ingrid S. Johnsrude, and Matthew H. Davis. 2010. ‘Hierarchical Processing for Speech in Human Auditory Cortex and Beyond’. *Frontiers in Human Neuroscience* 4(Article 51):3. doi: 10.3389/fnhum.2010.00051.

Peelle, Jonathan E., and Arthur Wingfield. 2016. ‘Listening Effort in Age-Related Hearing Loss’: *The Hearing Journal* 69(11):10–12. doi: 10.1097/01.HJ.0000508368.12042.08.

Perea Pérez, Francisca, Douglas E. H. Hartley, Pádraig T. Kitterick, and Ian M. Wiggins. 2022. ‘Perceived Listening Difficulties of Adult Cochlear-Implant Users Under Measures Introduced to Combat the Spread of COVID-19’. *Trends in Hearing* 26:1–22. doi: 10.1177/23312165221087011.

Perreau, Ann E., Yu-Hsiang Wu, Bailey Tatge, Diana Irwin, and Daniel Corts. 2017. ‘Listening Effort Measured in Adults with Normal Hearing and Cochlear Implants’. *Journal of the American Academy of Audiology* 28(8):685–97. doi: 10.3766/jaaa.16014.

Petersen, B., A. Gjedde, M. Wallentin, and P. Vuust. 2013. ‘Cortical Plasticity after Cochlear Implantation’. *Neural Plasticity* 2013:1–11. doi: 10.1155/2013/318521.

Phillips, Natalie A. 2016. ‘The Implications of Cognitive Aging for Listening and the Framework for Understanding Effortful Listening (FUEL)’. *Ear & Hearing* 37(1):44S-51S. doi: 10.1097/AUD.0000000000000309.

Phillips, Stephen. 2020. ‘Working through the Pandemic: Accelerating the Transition to Remote Working’. *Business Information Review* 37(3):129–34. doi: 10.1177/0266382120953087.

Pichora-Fuller, Kathleen, Sophia E. Kramer, Mark A. Eckert, Brent Edwards, BW Hornsby, Larry E. Humes, Ulrike Lemke, Thomas Lunner, Mohan Matthen, Carol L. Mackersie, Graham Naylor, Natalie A. Phillips, Michael Richter, Mary Rudner, Mitchell S. Sommers, Kelly L. Tremblay, and Arthur Wingfield. 2016. ‘Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL)’. *Ear and Hearing* 37(1):5S-27S. doi: 10.1097/AUD.0000000000000312.

Pichora-Fuller, Kathleen, Paul Mick, and Marilyn Reed. 2015. ‘Hearing, Cognition, and Healthy Aging: Social and Public Health Implications of the Links between Age-Related Declines in Hearing and Cognition’. *Seminars in Hearing* 36(03):122–39. doi: 10.1055/s-0035-1555116.

Pichora-Fuller, Kathleen, and Gurjit Singh. 2006. ‘Effects of Age on Auditory and Cognitive Processing: Implications for Hearing Aid Fitting and Audiologic Rehabilitation.’ *Trends in Amplification* 10(1):29–59. doi: 10.1177/108471380601000103.

Picou, Erin M., Todd A. Ricketts, and Benjamin W. Y. Hornsby. 2011. ‘Visual Cues and Listening Effort: Individual Variability’. *Journal of Speech, Language, and Hearing Research* 54(5):1416–30. doi: 10.1044/1092-4388(2011/10-0154).

Picou, Erin M., Todd A. Ricketts, and Benjamin W. Y. Hornsby. 2013. ‘How Hearing Aids, Background Noise, and Visual Cues Influence Objective Listening Effort’: *Ear and Hearing* 34(5):e52–64. doi: 10.1097/AUD.0b013e31827f0431.

Pinti, Paola, Ilias Tachtsidis, Antonia Hamilton, Joy Hirsch, Clarisse Aichelburg, Sam Gilbert, and Paul W. Burgess. 2018. ‘The Present and Future Use of Functional Near-Infrared Spectroscopy (FNIRS) for Cognitive Neuroscience: Advances in Using FNIRS in Cognitive Neuroscience’. *Annals of the New York Academy of Sciences*. doi: 10.1111/nyas.13948.

Piquado, Tepring, Derek Isaacowitz, and Arthur Wingfield. 2010. ‘Pupillometry as a Measure of Cognitive Effort in Younger and Older Adults’. *Psychophysiology* 47(3):560–69. doi: 10.1111/j.1469-8986.2009.00947.x.

Plichta, M. M., S. Heinzel, A. C. Ehlis, P. Pauli, and A. J. Fallgatter. 2007. ‘Model-Based Analysis of Rapid Event-Related Functional near-Infrared Spectroscopy (NIRS) Data: A Parametric Validation Study’. *NeuroImage* 35(2):625–34. doi: 10.1016/j.neuroimage.2006.11.028.

Plomp, R., and A. M. Mimpen. 1979. ‘Improving the Reliability of Testing the Speech Reception Threshold for Sentences.’ *Audiology : Official Organ of the International Society of Audiology* 18(1):43–52. doi: 10.3109/00206097909072618.

Poldrack, Russell A., Elise Temple, Athanassios Protopapas, Srikantan Nagarajan, Paula Tallal, Michael Merzenich, and John D. E. Gabrieli. 2001. ‘Relations between the Neural

Bases of Dynamic Auditory Processing and Phonological Processing: Evidence from FMRI’. *Journal of Cognitive Neuroscience* 13(5):687–97. doi: 10.1162/089892901750363235.

Pollonini, Luca, Cristen Olds, Homer Abaya, Heather Bortfeld, Michael S. Beauchamp, and John S. Oghalai. 2014. ‘Auditory Cortex Activation to Natural Speech and Simulated Cochlear Implant Speech Measured with Functional Near-Infrared Spectroscopy’. *Hearing Research* 309:84–93. doi: 10.1016/j.heares.2013.11.007.

Prodi, Nicola, and Chiara Visentin. 2015. ‘Listening Efficiency during Lessons under Various Types of Noise’. *The Journal of the Acoustical Society of America* 138(4):2438–48. doi: 10.1121/1.4932053.

Prodi, Nicola, Chiara Visentin, and Andrea Farnetani. 2010. ‘Intelligibility, Listening Difficulty and Listening Efficiency in Auralized Classrooms’. *The Journal of the Acoustical Society of America* 128(1):172–81. doi: 10.1121/1.3436563.

Pronk, Marieke, Dorly J. H. Deeg, Cas Smits, Theo G. van Tilburg, Dirk J. Kuik, Joost M. Festen, and Sophia E. Kramer. 2011. ‘Prospective Effects of Hearing Status on Loneliness and Depression in Older Persons: Identification of Subgroups’. *International Journal of Audiology* 50(12):887–96. doi: 10.3109/14992027.2011.599871.

Ptak, Radek. 2012. ‘The Frontoparietal Attention Network of the Human Brain: Action, Saliency, and a Priority Map of the Environment’. *The Neuroscientist* 18(5):502–15. doi: 10.1177/1073858411409051.

Quaresima, Valentina, Silvia Bisconti, and Marco Ferrari. 2012. ‘A Brief Review on the Use of Functional Near-Infrared Spectroscopy (FNIRS) for Language Imaging Studies in Human Newborns and Adults’. *Brain and Language* 121(2):79–89. doi: 10.1016/j.bandl.2011.03.009.

Rahne, Torsten, Laura Fröhlich, Stefan Plontke, and Luise Wagner. 2021. ‘Influence of Surgical and N95 Face Masks on Speech Perception and Listening Effort in Noise’. *PLOS ONE* 16(7):e0253874. doi: 10.1371/journal.pone.0253874.

Raine, Ch. 2014. ‘Access to and Uptake of Cochlear Implants in the UK’. *Ent and Audiology News*.

Rapport, Frances, Sarah E. Hughes, Isabelle Boisvert, Catherine M. McMahon, Jeffrey Braithwaite, Mona Faris, and Mia Bierbaum. 2020. ‘Adults’ Cochlear Implant Journeys through Care: A Qualitative Study’. *BMC Health Services Research* 20(1):457. doi: 10.1186/s12913-020-05334-y.

Razai, Mohammad S., Pippa Oakeshott, Hadyn Kankam, Sandro Galea, and Helen Stokes-Lampard. 2020. ‘Mitigating the Psychological Effects of Social Isolation during the Covid-19 Pandemic’. *BMJ* 369. doi: 10.1136/bmj.m1904.

Reay, Rebecca E., Jeffrey CL Looi, and Philip Keightley. 2020. ‘Telehealth Mental Health Services during COVID-19: Summary of Evidence and Clinical Practice’. *Australasian Psychiatry*. doi: 10.1177/1039856220943032.

Reed, Nicholas S., Lauren E. Ferrante, and Esther S. Oh. 2020. ‘Addressing Hearing Loss to Improve Communication During the COVID-19 Pandemic’. *Journal of the American Geriatrics Society* 68(9):1924–26. doi: 10.1111/jgs.16674.

Reilly, Jamie, Alexandra Kelly, Seung Hwan Kim, Savannah Jett, and Bonnie Zuckerman. 2019. ‘The Human Task-Evoked Pupillary Response Function Is Linear: Implications for Baseline Response Scaling in Pupillometry’. *Behavior Research Methods* 51(2):865–78. doi: 10.3758/s13428-018-1134-4.

Reimer, Jacob, Matthew J. McGinley, Yang Liu, Charles Rodenkirch, Qi Wang, David A. McCormick, and Andreas S. Tolias. 2016. ‘Pupil Fluctuations Track Rapid Changes in Adrenergic and Cholinergic Activity in Cortex’. *Nature Communications* 7:13289–13289. doi: 10.1038/ncomms13289.

Revelle, William. 2020. *Package ‘Psych’: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois.

Ribeiro, Vanessa Veis, Ana Paula Dassie-Leite, Eliane Cristina Pereira, Allicia Diely Nunes Santos, Perla Martins, and Roxane de Alencar Irineu. 2020. ‘Effect of Wearing a Face Mask on Vocal Self-Perception during a Pandemic’. *Journal of Voice* S0892199720303568. doi: 10.1016/j.jvoice.2020.09.006.

Richter, M., G. H. E. Gendolla, and R. A. Wright. 2016. ‘Three Decades of Research on Motivational Intensity Theory’. Pp. 149–86 in *Advances in Motivation Science*. Vol. 3. Elsevier.

RNID. 2022. ‘How Video Conferencing Apps Compare for Accessibility’. *RNID*. Retrieved 25 January 2022 (https://rnid.org.uk/information-and-support/technology-and-products/how-video-conferencing-apps-compare-for-accessibility/).

Robert, Christian P., ed. 2007a. ‘A Defense of the Bayesian Choice’. Pp. 507–30 in *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. New York, NY: Springer New York.

Robert, Christian P., ed. 2007b. ‘Hierarchical and Empirical Bayes Extensions’. Pp. 457–506 in *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. New York, NY: Springer New York.

Rødvik, Arne Kirkhorn, Janne von Koss Torkildsen, Ona Bø Wie, Marit Aarvaag Storaker, and Juha Tapio Silvola. 2018. ‘Consonant and Vowel Identification in Cochlear Implant Users Measured by Nonsense Words: A Systematic Review and Meta-Analysis’. *Journal of Speech, Language, and Hearing Research* 61(4):1023–50. doi: 10.1044/2018_JSLHR-H-16-0463.

Rönnberg, Jerker, Thomas Lunner, Adriana Zekveld, Patrik Sörqvist, Henrik Danielsson, Björn Lyxell, Örjan Dahlström, Carine Signoret, Stefan Stenfelt, M. Kathleen Pichora-Fuller, and Mary Rudner. 2013. ‘The Ease of Language Understanding (ELU) Model: Theoretical, Empirical, and Clinical Advances’. *Frontiers in Systems Neuroscience* 7. doi: 10.3389/fnsys.2013.00031.

Rönnberg, N., S. Stenfelt, and M. Rudner. 2011. ‘Testing Listening Effort for Speech Comprehension Using the Individuals’ Cognitive Spare Capacity’. *Audiology Research* 1(1S). doi: 10.4081/audiores.2011.e22.

Rosemann, Stephanie, and Christiane M. Thiel. 2020. ‘Neuroanatomical Changes Associated with Age-Related Hearing Loss and Listening Effort’. *Brain Structure and Function*. doi: 10.1007/s00429-020-02148-w.

Rovetti, Joseph, Huiwen Goy, M. Kathleen Pichora-Fuller, and Frank A. Russo. 2019. ‘Functional Near-Infrared Spectroscopy as a Measure of Listening Effort in Older Adults Who Use Hearing Aids’. *Trends in Hearing* 23:233121651988672. doi: 10.1177/2331216519886722.

Rowland, Stephen C., Douglas E. H. Hartley, and Ian M. Wiggins. 2018. ‘Listening in Naturalistic Scenes: What Can Functional Near-Infrared Spectroscopy and Intersubject Correlation Analysis Tell Us About the Underlying Brain Activity?’ *Trends in Hearing* 22:233121651880411. doi: 10.1177/2331216518804116.

Rudner, Mary. 2016. ‘Cognitive Spare Capacity as an Index of Listening Effort’: *Ear and Hearing* 37:69S-76S. doi: 10.1097/AUD.0000000000000302.

Rudner, Mary, Catharina Foo, Jerker Ronnberg, and Thomas Lunner. 2009. ‘Cognition and Aided Speech Recognition in Noise: Specific Role for Cognitive Factors Following Nine-Week Experience with Adjusted Compression Settings in Hearing Aids.’ *Scandinavian Journal of Psychology* 50(5):405–18. doi: 10.1111/j.1467-9450.2009.00745.x.

Rudner, Mary, Catharina Foo, Elisabet Sundewall-Thoren, Thomas Lunner, and Jerker Ronnberg. 2008. ‘Phonological Mismatch and Explicit Cognitive Processing in a Sample of 102 Hearing-Aid Users.’ *International Journal of Audiology* 47 Suppl 2:S91-98. doi: 10.1080/14992020802304393.

Rudner, Mary, Thomas Lunner, Thomas Behrens, Elisabet Sundewall Thorén, and Jerker Rönnberg. 2012. ‘Working Memory Capacity May Influence Perceived Effort during Aided Speech Recognition in Noise’. *Journal of the American Academy of Audiology* 23(8):577–89. doi: 10.3766/jaaa.23.7.7.

Russo, Francesca Yoshie, Michel Hoen, Chadlia Karoui, Thomas Demarcy, Marine Ardoint, Maria-Pia Tuset, Daniele De Seta, Olivier Sterkers, Ghizlène Lahlou, and Isabelle Mosnier. 2020. ‘Pupillometry Assessment of Speech Recognition and Listening Experience in Adult Cochlear Implant Patients’. *Frontiers in Neuroscience* 14. doi: 10.3389/fnins.2020.556675.

Saliba, Joe, Heather Bortfeld, Daniel J. Levitin, and John S. Oghalai. 2016. ‘Functional Near-Infrared Spectroscopy for Neuroimaging in Cochlear Implant Recipients’. *Hearing Research* 338:64–75. doi: 10.1016/j.heares.2016.02.005.

Salthouse, Timothy A., and David J. Madden. 2008. ‘Information Processing Speed and Aging.’ Pp. 221–41 in *Information processing speed in clinical populations., Studies on neuropsychology, neurology and cognition.* Philadelphia, PA, US: Taylor & Francis.

Sarampalis, Anastasios, Sridhar Kalluri, Brent Edwards, and Ervin Hafter. 2009. ‘Objective Measures of Listening Effort: Effects of Background Noise and Noise Reduction’. *Journal of Speech, Language, and Hearing Research* 52(5):1230–40. doi: 10.1044/1092-4388(2009/08-0111).

Saunders, Gabrielle H., Iain R. Jackson, and Anisa S. Visram. 2020. ‘Impacts of Face Coverings on Communication: An Indirect Impact of COVID-19’. *International Journal of Audiology* 1–12. doi: 10.1080/14992027.2020.1851401.

Saunders, Gabrielle H., and Francesca Oliver. 2022. ‘Impact of Hearing Loss on Communication During Remote Health Care Encounters’. *Telemedicine and E-Health*. doi: 10.1089/tmj.2021.0490.

Saunders, Gabrielle H., and Amber Roughley. 2021. ‘Audiology in the Time of COVID-19: Practices and Opinions of Audiologists in the UK’. *International Journal of Audiology* 60(4):255–62. doi: 10.1080/14992027.2020.1814432.

Schoenenberg, Katrin, Alexander Raake, and Judith Koeppe. 2014. ‘Why Are You so Slow? – Misattribution of Transmission Delay to Attributes of the Conversation Partner at the Far-End’. *International Journal of Human-Computer Studies* 72(5):477–87. doi: 10.1016/j.ijhcs.2014.02.004.

Scholkmann, Felix, Stefan Kleiser, Andreas Jaakko Metz, Raphael Zimmermann, Juan Mata Pavia, Ursula Wolf, and Martin Wolf. 2014. ‘A Review on Continuous Wave Functional Near-Infrared Spectroscopy and Imaging Instrumentation and Methodology’. *NeuroImage* 85:6–27. doi: 10.1016/j.neuroimage.2013.05.004.

Scott, S. K. 2000. ‘Identification of a Pathway for Intelligible Speech in the Left Temporal Lobe’. *Brain* 123(12):2400–2406. doi: 10.1093/brain/123.12.2400.

Sharma, Deepika, and Rangasayee Raghunath Rao. 2017. ‘The Effect of Captioning in Understanding Televised Speech in Geriatric Hearing Aid Users’. *Journal of Gerontology & Geriatric Research* 06(06). doi: 10.4172/2167-7182.1000459.

Sheldon, Signy, M. Kathleen Pichora-Fuller, and Bruce A. Schneider. 2008. ‘Priming and Sentence Context Support Listening to Noise-Vocoded Speech by Younger and Older Adults.’ *The Journal of the Acoustical Society of America* 123(1):489–99. doi: 10.1121/1.2783762.

Sherafati, Arefeh, Noel Dwyer, Aahana Bajracharya, Mahlega S. Hassanpour, Adam T. Eggebrecht, Jill B. Firszt, Joseph P. Culver, and Jonathan E. Peelle. 2021. ‘Dorsolateral Prefrontal Cortex Supports Speech Perception in Listeners with Cochlear Implants’. doi: 10.1101/2021.10.16.464654.

Sherafati, Arefeh, Noel Dwyer, Aahana Bajracharya, Mahlega Samira Hassanpour, Adam T. Eggebrecht, Jill B. Firszt, Joseph P. Culver, and Jonathan E. Peelle. 2022. ‘Prefrontal Cortex Supports Speech Perception in Listeners with Cochlear Implants’. *ELife* 11:e75323. doi: 10.7554/eLife.75323.

Shields, Callum, Mark Sladen, Iain Alexander Bruce, Karolina Kluk, and Jaya Nichani. 2023. ‘Exploring the Correlations Between Measures of Listening Effort in Adults and Children: A Systematic Review with Narrative Synthesis’. *Trends in Hearing* 27:233121652211371. doi: 10.1177/23312165221137116.

Shukla, Aishwarya, Michael Harper, Emily Pedersen, Adele Goman, Jonathan J. Suen, Carrie Price, Jeremy Applebaum, Matthew Hoyer, Frank R. Lin, and Nicholas S. Reed. 2020. ‘Hearing Loss, Loneliness, and Social Isolation: A Systematic Review’. *Otolaryngology--Head and Neck Surgery : Official Journal of American Academy of*

Otolaryngology-Head and Neck Surgery 162(5):622–33. doi: 10.1177/0194599820910377.

Singh, Archana K., and Ippeita Dan. 2006. ‘Exploring the False Discovery Rate in Multichannel NIRS’. *NeuroImage* 33(2):542–49. doi: 10.1016/j.neuroimage.2006.06.047.

Skidmore, Jeffrey A., Kara J. Vasil, Shuman He, and Aaron C. Moberly. 2020. ‘Explaining Speech Recognition and Quality of Life Outcomes in Adult Cochlear Implant Users: Complementary Contributions of Demographic, Sensory, and Cognitive Factors.’ *Otology & Neurotology : Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology* 41(7):e795–803. doi: 10.1097/MAO.0000000000002682.

Smeds, Karolina, Florian Wolters, and Martin Rung. 2015. ‘Estimation of Signal-to-Noise Ratios in Realistic Sound Scenarios’. *Journal of the American Academy of Audiology* 26(2):183–96. doi: 10.3766/jaaa.26.2.7.

Son, Joon Son, and Andrew Zisserman. 2017. ‘Lip Reading in Profile’. P. 155 in *Procedings of the British Machine Vision Conference 2017*. London, UK: British Machine Vision Association.

Stafford, Tom, Angelo Pirrone, Mike Croucher, and Anna Krystalli. 2020. ‘Quantifying the Benefits of Using Decision Models with Response Time and Accuracy Data’. *Behavior Research Methods* 52(5):2142–55. doi: 10.3758/s13428-020-01372-w.

Steffener, Jason, Matthias Tabert, Aaron Reuben, and Yaakov Stern. 2010. ‘Investigating Hemodynamic Response Variability at the Group Level Using Basis Functions’. *NeuroImage* 49(3):2113–22. doi: 10.1016/j.neuroimage.2009.11.014.

Steinhauer, Stuart R., Greg J. Siegle, Ruth Condray, and Misha Pless. 2004. ‘Sympathetic and Parasympathetic Innervation of Pupillary Dilation during Sustained Processing’. *International Journal of Psychophysiology* 52(1):77–86. doi: 10.1016/j.ijpsycho.2003.12.005.

Stenbäck, Victoria, Erik Marsja, Rachel Ellis, and Jerker Rönnberg. 2023. ‘Relationships between Behavioural and Self-Report Measures in Speech Recognition in Noise’. *International Journal of Audiology* 1–9. doi: 10.1080/14992027.2022.2047232.

Strand, Julia F., Violet A. Brown, Madeleine B. Merchant, Hunter E. Brown, and Julia Smith. 2018. ‘Measuring Listening Effort: Convergent Validity, Sensitivity, and Links With Cognitive and Personality Measures’. *Journal of Speech, Language, and Hearing Research* 61(6):1463–86. doi: 10.1044/2018_JSLHR-H-17-0257.

Strauss, Daniel J., and Alexander L. Francis. 2017. ‘Toward a Taxonomic Model of Attention in Effortful Listening’. *Cognitive, Affective, & Behavioral Neuroscience* 17(4):809–25. doi: 10.3758/s13415-017-0513-0.

Strelnikov, K., M. Marx, S. Lagleyre, B. Fraysse, O. Deguine, and P. Barone. 2015. ‘PET-Imaging of Brain Plasticity after Cochlear Implantation’. *Hearing Research* 322:180–87. doi: 10.1016/j.heares.2014.10.001.

Suda, Masashi, Yuichi Takei, Yoshiyuki Aoyama, Kosuke Narita, Toshimasa Sato, Masato Fukuda, and Masahiko Mikuni. 2010. ‘Frontopolar Activation during Face-to-Face

Conversation: An in Situ Study Using near-Infrared Spectroscopy.' *Neuropsychologia* 48(2):441–47. doi: 10.1016/j.neuropsychologia.2009.09.036.

Tagupa, Hyacinth. 2020. 'Social Isolation, Loneliness, and Hearing Loss During COVID-19'. *The Hearing Journal* 73(5):46–47. doi: 10.1097/01.HJ.0000666456.65020.b9.

Tavanai, Elham, Nematollah Rouhbakhsh, and Zohreh Roghani. 2021. 'A Review of the Challenges Facing People with Hearing Loss during the COVID-19 Outbreak: Toward the Understanding the Helpful Solutions'. *Auditory and Vestibular Research* 30(2):62–63. doi: 10.18502/avr.v30i2.6091.

Ten Hulzen, Richard D., and David A. Fabry. 2020. 'Impact of Hearing Loss and Universal Face Masking in the COVID-19 Era'. *Mayo Clinic Proceedings* 95(10):2069–72. doi: 10.1016/j.mayocp.2020.07.027.

Thammaiah, Spoorthi, Vinaya Manchaiah, Vijayalakshmi Easwar, Rajalakshmi Krishna, and Bradley McPherson. 2017. 'Psychometric Properties of the Hearing Handicap Questionnaire: A Kannada (South-Indian) Translation'. *International Journal of Audiology* 56(3):194–201. doi: 10.1080/14992027.2016.1247500.

Townsend, James T., and F. Gregory Ashby. 1983. 'The Stochastic Modeling of Elementary Psychological Processes'. *Cambridge: Cambridge University Press*. doi: https://doi.org/10.2307/1422636.

Tun, Patricia A., Sandra McCoy, and Arthur Wingfield. 2009. 'Aging, Hearing Acuity, and the Attentional Costs of Effortful Listening.' *Psychology and Aging* 24(3):761–66. doi: 10.1037/a0014802.

UCL. 2020. 'Remote Working Guidelines'. *UCL Deafness Cognition and Language Research Centre*. Retrieved 27 January 2022 (https://www.ucl.ac.uk/dcal/remote-working-guidelines).

Van Ravenzwaaij, Don, Pete Cassey, and Scott D. Brown. 2018. 'A Simple Introduction to Markov Chain Monte–Carlo Sampling'. *Psychonomic Bulletin & Review* 25(1):143–54. doi: 10.3758/s13423-016-1015-8.

Vandierendonck, André. 2017. 'A Comparison of Methods to Combine Speed and Accuracy Measures of Performance: A Rejoinder on the Binning Procedure'. *Behavior Research Methods* 49(2):653–73. doi: 10.3758/s13428-016-0721-5.

Ventry, Ira M., and Barbara E. Weinstein. 1982. 'The Hearing Handicap Inventory for the Elderly: A New Tool'. 128–34. doi: 10.1097/00003446-198205000-00006.

Visentin, Chiara, Nicola Prodi, Simone Torresin, Francesca Cappelletti, and Andrea Gasparella. 2017. 'Listening Efficiency in University Classrooms: A Comparison between Native and Non-Native Listeners'. 12.

Visentin, Chiara, Chiara Valzolgher, Matteo Pellegatti, Paola Potente, Francesco Pavani, and Nicola Prodi. 2022. 'A Comparison of Simultaneously-Obtained Measures of Listening Effort: Pupil Dilation, Verbal Response Time and Self-Rating.' *International Journal of Audiology* 1–13. doi: 10.1080/14992027.2021.1921290.

Walsh, B., F. Tian, J. A. Tourville, M. A. Yücel, T. Kuczek, and A. J. Bostian. 2017. ‘Hemodynamics of Speech Production: An FNIRS Investigation of Children Who Stutter’. *Scientific Reports* 7(1):4034. doi: 10.1038/s41598-017-04357-6.

Wang, Bin, Yukun Liu, Jing Qian, and Sharon K. Parker. 2021. ‘Achieving Effective Remote Working During the COVID-19 Pandemic: A Work Design Perspective’. *Applied Psychology* 70(1):16–59. doi: https://doi.org/10.1111/apps.12290.

Wang, Cuiyan, Riyu Pan, Xiaoyang Wan, Yilin Tan, Linkang Xu, Roger S. McIntyre, Faith N. Choo, Bach Tran, Roger Ho, Vijay K. Sharma, and Cyrus Ho. 2020. ‘A Longitudinal Study on the Mental Health of General Population during the COVID-19 Epidemic in China’. *Brain, Behavior, and Immunity* 87:40–48. doi: 10.1016/j.bbi.2020.04.028.

Wang, Yang, Graham Naylor, Sophia E. Kramer, Adriana A. Zekveld, Dorothea Wendt, Barbara Ohlenforst, and Thomas Lunner. 2018. ‘Relations Between Self-Reported Daily-Life Fatigue, Hearing Status, and Pupil Dilation During a Speech Perception in Noise Task’: *Ear and Hearing* 39(3):573–82. doi: 10.1097/AUD.0000000000000512.

Wendling, Jean-Michel, Thibaut Fabacher, Philippe-Pierre Pébaÿ, Isabelle Cosperec, and Michaël Rochoy. 2021. ‘Experimental Efficacy of the Face Shield and the Mask against Emitted and Potentially Received Particles’. *International Journal of Environmental Research and Public Health* 18(4):1942. doi: 10.3390/ijerph18041942.

Wendt, Dorothea, Torsten Dau, and Jens Hjortkjær. 2016. ‘Impact of Background Noise and Sentence Complexity on Processing Demands during Sentence Comprehension’. *Frontiers in Psychology* 7:345–345. doi: 10.3389/fpsyg.2016.00345.

White, Bradley E., and Clifton Langdon. 2021. ‘The Cortical Organization of Listening Effort: New Insight from Functional near-Infrared Spectroscopy’. *NeuroImage* 240:118324. doi: 10.1016/j.neuroimage.2021.118324.

White, Corey N., Ryan A. Curl, and Jennifer F. Sloane. 2016. ‘Using Decision Models to Enhance Investigations of Individual Differences in Cognitive Neuroscience’. *Frontiers in Psychology* 7. doi: 10.3389/fpsyg.2016.00081.

White, Sarah J., Serena Barello, Eugenia Cao di San Marco, Cinzia Colombo, Edgard Eeckman, Conor Gilligan, Guendalina Graffigna, T. Jirasevijinda, Paola Mosconi, Judy Mullan, Shakaib Ur Rehman, Sara Rubinelli, Elena Vegni, and Demi Krystallidou. 2021. ‘Critical Observations on and Suggested Ways Forward for Healthcare Communication during COVID-19: PEACH Position Paper’. *Patient Education and Counseling* 104(2):217–22. doi: 10.1016/j.pec.2020.12.025.

WHO. 2020a. *Advice on the Use of Masks in the Context of COVID-19: Interim Guidance, 6 April 2020*. Geneva: World Health Organization.

WHO. 2020b. *Mental Health and Psychosocial Considerations during the COVID-19 Outbreak*. World Health Organization.

Wie, Ona B., Are Hugo Pripp, and Ole Tvete. 2010. ‘Unilateral Deafness in Adults: Effects on Communication and Social Interaction.’ *The Annals of Otology, Rhinology, and Laryngology* 119(11):772–81.

Wiggins, Ian M., Carly A. Anderson, Pádraig T. Kitterick, and Douglas E. H. Hartley. 2016. 'Speech-Evoked Activation in Adult Temporal Cortex Measured Using Functional near-Infrared Spectroscopy (FNIRS): Are the Measurements Reliable?' *Hearing Research* 339:142–54. doi: 10.1016/j.heares.2016.07.007.

Wijayasiri, Pramudi, Douglas E. H. Hartley, and Ian M. Wiggins. 2017. 'Brain Activity Underlying the Recovery of Meaning from Degraded Speech: A Functional near-Infrared Spectroscopy (FNIRS) Study'. *Hearing Research* 351:55–67. doi: 10.1016/j.heares.2017.05.010.

Wild, C. J., A. Yusuf, D. E. Wilson, J. E. Peelle, M. H. Davis, and I. S. Johnsrude. 2012. 'Effortful Listening: The Processing of Degraded Speech Depends Critically on Attention'. *Journal of Neuroscience* 32(40):14010–21. doi: 10.1523/JNEUROSCI.1528-12.2012.

Willis, Helen. 2018. 'The Feasibility of the Dual-Task Paradigm as a Framework for a Clinical Test of Listening Effort in Cochlear Implant Users'.

Wilson, Blake S., and Michael F. Dorman. 2008. 'Cochlear Implants: A Remarkable Past and a Brilliant Future'. *Hearing Research* 242(1–2):3–21. doi: 10.1016/j.heares.2008.06.005.

Wilson, Harper L., Jacob Crouch, Marissa Schuh, Jennifer Shinn, and Matthew L. Bush. 2021. 'Impacts of the COVID-19 Pandemic on Communication and Healthcare Access for Adults With Hearing Loss'. *Otology & Neurotology* 42(8):1156–64. doi: 10.1097/MAO.0000000000003203.

Wilson, Richard H., Rachel A. McArdle, and Sherri L. Smith. 2007. 'An Evaluation of the BKB-SIN, HINT, QuickSIN, and WIN Materials on Listeners With Normal Hearing and Listeners With Hearing Loss'. *Journal of Speech, Language, and Hearing Research* 50(4):844–56. doi: 10.1044/1092-4388(2007/059).

Winn, Matthew B., Jan R. Edwards, and Ruth Y. Litovsky. 2015. 'The Impact of Auditory Spectral Resolution on Listening Effort Revealed by Pupil Dilation': *Ear and Hearing* 36(4):e153–65. doi: 10.1097/AUD.0000000000000145.

Winn, Matthew B., and Katherine H. Teece. 2020. 'Slower Speaking Rate Reduces Listening Effort Among Listeners With Cochlear Implants'. *Ear & Hearing* Publish Ahead of Print. doi: 10.1097/AUD.0000000000000958.

Winn, Matthew B., and Katherine H. Teece. 2021. 'Listening Effort Is Not the Same as Speech Intelligibility Score'. *Trends in Hearing* 25:233121652110276. doi: 10.1177/23312165211027688.

Winn, Matthew B., and Katherine H. Teece. 2022. 'Effortful Listening Despite Correct Responses: The Cost of Mental Repair in Sentence Recognition by Listeners With Cochlear Implants'. *Journal of Speech, Language, and Hearing Research* 1–15. doi: 10.1044/2022_JSLHR-21-00631.

Winn, Matthew B., Dorothea Wendt, Thomas Koelewijn, and Stefanie E. Kuchinsky. 2018. 'Best Practices and Advice for Using Pupillometry to Measure Listening Effort: An Introduction for Those Who Want to Get Started'. *Trends in Hearing* 22:233121651880086. doi: 10.1177/2331216518800869.

Wolf, Martin, Ursula Wolf, Vlad Toronov, Antonios Michalos, L. Adelina Paunescu, Jee Hyun Choi, and Enrico Gratton. 2002. 'Different Time Evolution of Oxyhemoglobin and Deoxyhemoglobin Concentration Changes in the Visual and Motor Cortices during Functional Stimulation: A Near-Infrared Spectroscopy Study'. *NeuroImage* 16(3):704–12. doi: 10.1006/nimg.2002.1128.

Woltz, Dan J., and Christopher A. Was. 2006. 'Availability of Related Long-Term Memory during and after Attention Focus in Working Memory'. *Memory & Cognition* 34(3):668–84. doi: 10.3758/BF03193587.

Wu, Yu-Hsiang, Elizabeth Stangl, Octav Chipara, Syed Shabih Hasan, Anne Welhaven, and Jacob Oleson. 2018. 'Characteristics of Real-World Signal to Noise Ratios and Speech Listening Situations of Older Adults With Mild to Moderate Hearing Loss': *Ear and Hearing* 39(2):293–304. doi: 10.1097/AUD.0000000000000486.

Xiu, Bowen, Brandon T. Paul, Joseph M. Chen, Trung N. Le, Vincent Y. Lin, and Andrew Dimitrijevic. 2022. 'Neural Responses to Naturalistic Audiovisual Speech Are Related to Listening Demand in Cochlear Implant Users'. *Frontiers in Human Neuroscience* 16:1043499. doi: 10.3389/fnhum.2022.1043499.

Yamada, Toru, Shinji Umeyama, and Keiji Matsuda. 2012. 'Separation of FNIRS Signals into Functional and Systemic Components Based on Differences in Hemodynamic Modalities' edited by J.-C. Baron. *PLoS ONE* 7(11):e50271. doi: 10.1371/journal.pone.0050271.

Yerkes, Robert M., and John D. Dodson. 1908. 'The Relation of Strength of Stimulus to Rapidity of Habit-Formation'. *Journal of Comparative Neurology and Psychology* 18(5):459–82. doi: 10.1002/cne.920180503.

Zekveld, Adriana A., Erwin L. J. George, Sophia E. Kramer, S. Theo Goverts, and Tammo Houtgast. 2007. 'The Development of the Text Reception Threshold Test: A Visual Analogue of the Speech Reception Threshold Test'. *Journal of Speech, Language, and Hearing Research* 50(3):576–84. doi: 10.1044/1092-4388(2007/040).

Zekveld, Adriana A., Dirk J. Heslenfeld, Joost M. Festen, and Ruurd Schoonhoven. 2006. 'Top–down and Bottom–up Processes in Speech Comprehension'. *NeuroImage* 32(4):1826–36. doi: 10.1016/j.neuroimage.2006.04.199.

Zekveld, Adriana A., Dirk J. Heslenfeld, Ingrid S. Johnsrude, Niek J. Versfeld, and Sophia E. Kramer. 2014. 'The Eye as a Window to the Listening Brain: Neural Correlates of Pupil Size as a Measure of Cognitive Listening Load'. *NeuroImage* 101:76–86. doi: 10.1016/j.neuroimage.2014.06.069.

Zekveld, Adriana A., Thomas Koelewijn, and Sophia E. Kramer. 2018. 'The Pupil Dilation Response to Auditory Stimuli: Current State of Knowledge'. *Trends in Hearing* 22:233121651877717. doi: 10.1177/2331216518777174.

Zekveld, Adriana A., and Sophia E. Kramer. 2014. 'Cognitive Processing Load across a Wide Range of Listening Conditions: Insights from Pupillometry: Processing Load across a Wide Range of Listening Conditions'. *Psychophysiology* 51(3):277–84. doi: 10.1111/psyp.12151.

Zekveld, Adriana A., Sophia E. Kramer, and Joost M. Festen. 2011. ‘Cognitive Load During Speech Perception in Noise: The Influence of Age, Hearing Loss, and Cognition on the Pupil Response’: *Ear and Hearing* 32(4):498–510. doi: 10.1097/AUD.0b013e31820512bb.

Zhang, Yan, Xin Liu, Dan Liu, Chunling Yang, Qisong Wang, Jinwei Sun, and Kuanquan Wang. 2018. ‘Evoked Hemodynamic Response Estimation to Auditory Stimulus Using Recursive Least Squares Adaptive Filtering with Multidistance Measurement of Near-Infrared Spectroscopy’. *Journal of Healthcare Engineering* 2018:1–8. doi: 10.1155/2018/7609713.

Zhou, Xin, Abd-Krim Seghouane, Adnan Shah, Hamish Innes-Brown, Will Cross, Ruth Litovsky, and Colette M. McKay. 2018. ‘Cortical Speech Processing in Postlingually Deaf Adult Cochlear Implant Users, as Revealed by Functional Near-Infrared Spectroscopy’. *Trends in Hearing* 22:233121651878685. doi: 10.1177/2331216518786850.