Valuing hearing-related changes in health-related quality of life.

NATALIE JADE WILLIAMS, BSc. MSc.

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

JULY 2019

Abstract

Within healthcare systems different areas of healthcare compete for limited resources. Allocating resources across healthcare programmes efficiently requires accurate, verifiable, robust methods. To ensure optimum resource allocations health conditions must be evaluated, taking into account their effects on quality of life, and the cost of receiving treatments. However, before a health state can be evaluated it must first be described to the individual valuing it, in order for an individual to know exactly what they are evaluating. With hearing loss, as with many other medical conditions, it is difficult to describe the impact of having a hearing loss on quality of life, either using general health dimensions or with words alone. This research aims to explore alternative methods of describing hearing-related heath states that have been prompted by recent advances in health valuation.

Nine experiments were carried out in order to achieve this goal. Experiment 1 aimed to validate the use of acoustic simulations to describe health states related to hearing and Experiment 2 validated their ability to recreate the listening difficulties reported by those with single-sided deafness (SSD). After alterations to the simulations Experiment 3 revalidated the ability of simulations to recreate the listening difficulties reported by those with SSD. Experiment 4 and 5 evaluated whether engagement with simulations is required and whether a simulation of no hearing loss should be provided as a comparator to hearing loss simulations. Experiment 6 assessed whether simulated hearing loss had differential effects depending on whether it supplemented with a detailed written description of the effects of that hearing loss on the health state. Experiment 7 examined whether acoustic simulations could convey the effects of an intervention, specifically the effect of cochlear implant. Experiment 8 assessed the feasibility of collecting health state valuation data on a large scale via a web-based interface. Finally, Experiment 9 aimed to identify other possible situations to holistically describe an SSD-related health state from a patient’s perspective.

Simulations of hearing loss appear to be a valid and reliable method for describing hearing-related health states, with broad potential to benefit other health states and areas such as education about hearing loss effects. Further research should explore these alternative uses so that this resource can be utilised to its full potential.

Acknowledgements

There are so many people I owe my thanks to for their help and support, and without them this thesis would not have been possible.

First and foremost, I would like to thank my parents for their never-ending encouragement and support throughout all my years of schooling, for teaching me valuable life lessons, a strong work ethic, patience, love and compassion. Mom and Dad without your love and positive influence in my life I could not be the woman I am today. I would like to thank my sister for teaching me patience, for reminding me to take time for myself, for the perfectly timed hugs and for loving and supporting me the way only a sister can. I would like to thank my partner for being there for me every day with words of reassurance, for cheering me on when the seemingly gargantuan task ahead of me all became too much for me to bear and for helping me every step of the way with anything I needed, be that proof reading or emergency snacks. You have been a pillar of support for me throughout this process; you were right it was all worth it in the end.

I would like to show my gratitude to NIHR for funding this project, and my supervision team: Padraig Kitterick, Robert Pierzycki, and Michael Akeroyd, for your guidance, support and time. I would like to thank Sandra Smith for helping me with the plethora of tasks I found on my to do list just to get my experiments underway, for advising me when I needed it the most, for boosting “gym” morale and for making myself, and many others, smile through the stressful times just by being your kind, caring and unique self. I would like to thank my friends for their encouragement and support throughout this long process, and finally, I would like to thank the “gymers”, who all became my firm friends, for the tea and biscuits, baked treats, pearls of wisdom, fool proof advice, and a good giggle when I needed it the most.

Declaration

I certify that this is my own work, except where indicated by referencing.

No part of this thesis has been submitted elsewhere for any other degree or qualification.

Natalie Williams Date

Table of Contents

Table of Tables

Table of Figures

List of Abbreviations

Abbreviation	Definition
A/ A alone	Audio alone
ANOVA	Analysis of Variance
ASH	Acoustic Simulations of Hearing
AT	Audio and Text
BKB	Bamford-Kowal-Bench
CA	Content analysis
CBA	Cost benefit analysis
CCA	Cost consequence analysis
CEA	Cost effectiveness analysis
CI	Cochlear Implant
CRM	Co-ordinate response measure
CUA	Cost utility analysis
EQ 5D	Euro-Qol 5 Dimension
FT	Full Text
FTA	Full Text and ASH
HRQoL	Hearing-related Quality of Life
HU	Health utility
HUI 3	Health Utilities Index Mark III
ICC	Intraclass Correlation Co-efficient
ICER	Incremental Cost-Effectiveness Ratio
NICE	National Institute for Health and Care Excellence.
QALY	Quality Adjusted Life Years
QoL	Quality of Life
SD	Standard deviation
SG	Standard Gamble
SiS	Speech in Speech
SiS+B	Speech in Speech with background noise
SNR	Signal-to-noise ratio

SSD	Single-Sided Deafness
T	Text alone
TTO	Time Trade Off
U VAS	VAS data transformed into utility values
VAS	Visual Analogue Scale
WHO	World Health Organisation

1 Introduction

Hearing loss is measured in terms of the type of loss: conductive or sensorineural hearing loss, and the severity of the hearing loss: ranging from mild to moderate to severe. A conductive hearing loss occurs when sounds cannot reach the inner ear, meaning that within either the ear canal, the ear drum, or the middle ear obstacles have occurred that prevent sound from reaching the inner ear. A sensorineural hearing loss is caused by changes to the inner ear and the first part of the hearing nerve. To combat a conductive hearing loss sound must be made louder to achieve a normal stimulation of the inner ear with interventions such as hearing aids and cochlear implants (Elberling & Worsen 2006).

Speech communication is an important element in everyday lives. Individuals with a hearing impaired may start to notice that it becomes more difficult to follow a conversation when many people are present or in noisy surroundings and that it is increasingly difficult to localise sound. Speech is an essential part of communication with other people and it is important for the development of social skills as well as for education (Elberling & Worsen 2006). According to the World Health Organisation [WHO] (2020) around 5% of the world's population has a disabling hearing loss, with 91% of those with a hearing loss being adults. The WHO highlights many financial and societal costs of hearing loss: health care costs, costs to the education sector for providing additional support to individuals with hearing loss, a loss of productivity within the workforce due to unemployment or premature retirement. The significant impact on everyday life should also be noted with hearing loss causing social isolation, feelings of loneliness and frustration, particularly among older people with hearing loss (WHO).

With such a large number of individuals experiencing hearing loss, there will inevitably be a need for funds to pay for treatments and services. In publicly funded healthcare systems (such as the National Health Service [NHS]) and in privately funded healthcare systems, different areas of healthcare compete for limited resources. Those charged with deciding where the resources will be allocated need to have accurate, verifiable, robust methods of allocating resources across healthcare programmes efficiently. The decision process must use information on both the costs of providing a service, and the health benefits generated by the service (Drummond et al., 2005).

Calculating the direct monetary cost of providing a healthcare service or intervention is a straightforward question of calculating the cost per treatment. However, in economic evaluations it is not only the direct costs of an intervention that are estimated, societal costs are often included, which can be challenging. It is also a challenge to calculate the units used to measure the benefits and effectiveness of a service or interventions vary considerably between interventions; e.g. additional years of life resulting from receiving a treatment for cancer or decibels of additional hearing sensitivity resulting from the use of a hearing aid.

This variability often prevents direct comparisons of different services in terms of their relative effectiveness being possible.

A solution to this problem is to measure the effectiveness of services and interventions in general terms, which are easily applicable and transferable to all areas of healthcare. The most common approach currently used to value health states and the effects of interventions is to use the Quality-Adjusted Life Year (QALY) method. This method uses both the duration, and the quality of life, to describe both the negative impact of disease on a patient's health and the benefits to health provided by a treatment or intervention. There are an immense range of possible instruments that can be used for measuring quality of life. However, Health Economists have advocated an approach in which informants value the quality of life of a particular state of health by positioning it on a scale bounded by the fixed states of death and optimal health (Gold et al., 1996).

The health state must first be described to the individual valuing it in order for an individual to know exactly what they are evaluating. A common approach to describing the health state has been to describe health states using a limited set of health dimensions, e.g. degrees of pain, mobility, depression, etc. (Brooks, 1996; Feeney et al., 2002). However, not all health dimensions are appropriate, or relevant, when describing certain health states, including hearing loss (Barton et al., 2004). An alternative method to using these dimensions is to describe the health state using short narratives, which outline the main features of the particular health state. These descriptions may also describe the limitations that the health state places on everyday social and vocational activities. These text-based descriptions have been found to be more sensitive in describing small changes in a health condition and more reflective of patient concerns (Brazier et al., 2007).

With hearing loss, as with many other medical conditions, it is difficult to describe the impact of having a hearing loss on quality of life, either using a set of general health dimensions or with words alone. It is not clear, therefore, whether valuations of health states associated with hearing loss, which have been described using health dimensions or text descriptions, could ever adequately reflect the impact of hearing loss on a patient's health-related quality of life. Consequently, alternative methods of describing health states may be necessary to ensure that the impacts of hearing loss, and the benefits of hearing services, are valued appropriately when conducting health economic assessments.

This research aims to explore alternative methods of describing hearing-related heath states that have been prompted by recent advances in health valuation. More recent approaches in other areas of health have employed additional forms of information in novel modalities, such as videos of patient interviews, to provide a more comprehensive description of a health state to the informants who are tasked with valuing it. What is relatively unique about hearing loss compared to many medical conditions is that it is possible to accurately simulate this loss of acoustic information for normal hearing individuals and in addition, the effects of interventions, such as cochlear implants, on hearing abilities.

1.2 Aims and objectives of this thesis

The purpose of this research is to assess the validity and reliability of using Acoustic Simulation of Hearing Loss (ASH) to obtain valuations of health states related to hearing loss. The specific aims of this study are to:

(i) to assess the validity of using ASH to describe a hearing-related health and to value the quality of life of health states related to hearing loss

(ii) to assess the reliability of quality of life valuations obtained using ASH

(iii) to explore how best to use and present ASH when collecting health state preference data for health states related to hearing loss

(iv) to explore what situations patients think should be simulated when using ASH to describe health states related to hearing loss

1.3 Summary of chapters

1.3.1 Chapter 2: Literature Review

This chapter reviews and discusses literature covering four key areas: economic evaluation, hearing loss, describing and measuring health, and methods of obtaining health state preferences (utility values). The *economic evaluation* section covers: the history, theories, concepts and methods of economic evaluation. The *hearing loss* section details: the aetiology of hearing loss, the aetiology of single-sided deafness (SSD, the hearing loss that this thesis focuses on), the treatment options available, and previous work to simulate and recreate hearing impairments. The *describing and measuring health* section examines: health valuation questionnaires, text-based descriptions of health, and the use of text-based descriptions in hearing research. The *methods of obtaining health state preferences* discusses: Standard Gamble, Time Trade Off, and Visual Analogue Scales.

1.3.2 Chapter 3: Feasibility of using acoustic simulations to describe health states

This chapter reports the development and pilot testing of the initial simulations and the subsequent steps taken to evaluate the potential of using of acoustic simulations to describe hearing-related health states. The simulation environment and the situation being simulated are described in detail. The pilot testing of the simulations is detailed. The subsequent experimental validation of both the use of acoustic simulations to describe health states related to hearing (Experiment 1) and their ability to recreate the listening difficulties reported by those with SSD (Experiment 2) is reported.

1.3.3 Chapter 4: Further development of acoustic simulations

This chapter describes the improvements made to the simulations based upon the results and feedback from the initial studies. The chapter reports pilot testing carried out to ascertain whether the changes made to the simulations addressed issues identified during previous experiments, and to identify optimum parameters for the simulations; e.g. signal to noise ratio (SNR). Experiment 3 is then reported which used a listening task to validate the effectiveness of the simulations in recreating the listening difficulties of those with SSD is reported.

1.3.4 Chapter 5: Refining the methodology of applying acoustic simulations

This chapter reports the results of Experiment 4 which aimed to evaluate whether active engagement with simulations is required compared to passive listening to simulations. Experiment 5 is also described, which aimed to determine whether a simulation of healthy hearing (i.e. of no hearing loss) should be provided as a comparator to the hearing loss simulation. The chapter also reports the results of Experiment 6 that aimed to assess whether simulated hearing loss had differential effects depending on whether it supplemented by a sparse or detailed written description of the effects of that hearing loss on the health state.

1.3.5 Chapter 6: Simulating interventions and the practicality of delivering acoustic simulations remotely

This chapter reports Experiment 7 that was designed to examine whether acoustic simulations could convey the effects of an intervention, specifically the effect of cochlear implantation for those with SSD. Experiment 8 is also reported, which aimed to assess whether it is practical to present acoustic simulations of hearing loss via an internet-based utility assessment tool as a way of increasing the feasibility of collecting health state valuation data on a large scale in future studies.

1.3.6 Chapter 7: Identifying supplementary situations to simulate with ASH

This chapter discusses the results of a qualitative focus group study which was conducted with adults who have SSD. The qualitative study aimed to identify other possible situations to simulate using ASH to describe health states related to SSD and to identify the set of acoustic simulations that would holistically describe an SSD-related health state from the patient’s perspective.

1.3.7 Chapter 8: Discussion

This chapter will present the main points for general discussion that have arisen from the work, summarise the findings of the thesis, and identify and discuss potential avenues for future research.

2 Literature Review

2.1 Economic evaluation

The finite nature of the resources that health care providers have available to them makes it necessary for decisions to be made about which heath care options should receive funding. Simply put economic evaluation assesses input and outputs of an activity or the costs and consequences of an activity. Without a way of measuring and comparing outputs, an improvement in quality of life and inputs, and the monetary cost of an intervention, there is no way to objectively and systematically judge the value for money of an intervention (Drummond et al., 2005). The aim of economic evaluations of health care is to support the decision makers in making justifiable and transparent funding decisions for health intervention (Brazier et al., 2007). The term ‘interventions’ in this review is broadly interpreted and can refer to new technologies, drugs, surgical procedures, non-invasive treatments, and so on.

In the early days of economic evaluation the practice was primarily an academic pursuit, with no real application of evaluation theories to health policies and decision making. Once the field of health economics, and the theories behind economic evaluation, were further developed and improved upon economic evaluation began to be applied to the evaluation of emerging medical procedures and devices. The first example of work evaluating the efficacy of a treatment route was an article by David Poole in 1969 (as cited in Buxton 2006). Poole looked at the economics of screening programmes for groups at high risk of contracting Tuberculosis and whether the screening programmes that targeted Tuberculosis were cost-effective.

From a UK perspective, a key moment in the economic evaluation of health care evaluation was the formation of the National Institute of Clinical Excellence (NICE; now National Institute of Health and Care Excellence) in 1999. The original aim of the organisation was to investigate the clinical effectiveness and cost-effectiveness of newly emerging drugs and treatments. In April 2013, NICE’s role changed as it became a Non-Departmental Public Body. Independent from the British government, NICE now also develops guidance and quality standards in social care for use by the NHS, local authorities, social care organisations, charities and any individual or organisation responsible for commissioning or providing healthcare services, public health services or social care services (NICE, 2014). In order to decide whether or not a treatment or technology is cost effective, NICE or those evaluating the new treatment will collect information about the health of individuals before and after the receiving the new treatment or healthcare technology. This process aims to quantify the benefit to overall health and well-being which is usually quantified in terms of an increase in overall quality of life. The outcome of the economic evaluation process, a comparison of the cost of a treatment compared to the health benefits it provides, is then used by the different trusts within the National Health Service (NHS) to inform decisions about the selection of new and existing treatments and devices (Brazier et al., 2007). Using NICE's economic evaluation data helps the decision makers within the NHS trusts, and private healthcare providers, to fund interventions that allow practitioners to deliver the best possible care, while also assuring that the care and treatments provided represent good value for money, with no inequalities and variations in the care provided.

2.2 Theories, concepts, and methods

The simplest purpose of performing an economic evaluation of a new piece of health technology or a new treatment option is to determine whether the benefit that people would gain (in terms of improved quality of life) from the new treatment or technology outweighs the direct monetary costs of funding the treatment and the benefit that is foregone through displacement of funding to other services and/or interventions. The field of Health Economics has developed several theories and methodologies that can be used to assess and compare the costs and the benefits of health care interventions and medical devices (Brazier et al., 2007). The techniques that will be discussed in this review are: cost effectiveness analysis (CEA), cost utility analysis (CUA), cost benefit analysis (CBA) and cost consequence analysis (CCA).

2.2.1 Cost effectiveness analysis (CEA)

CEA is a method that compares the cost of various treatment options and interventions that can be used to achieve a certain outcome. The two, or more, interventions that are being compared can be very similar, e.g. they could be two different drugs that both treat a specific illness, or they could starkly contrast one another, such as a surgical procedure versus a non-invasive screening process. CEA can be used to assess interventions in terms of directly measured units, such as, detection rate, reduction of illness or years of life saved. The health benefit estimations that are used as outcomes in CEA can be very uncertain, CEA does not question the worth of the tested outcome, i.e. it does not tell you whether the chosen outcome was the best or most appropriate one to use. Where two or more interventions are deemed to have similar benefit levels using the same outcome, e.g. life years saved, then the intervention with the lowest costs is the most cost-effective alternative. Equally, if both interventions have similar costs then the treatment with the greater benefits is most cost-effective. Where both costs and benefits differ, the CEA compares the size of the additional costs to the size of the additional benefits. This result is usually expressed in terms of a ratio of 'incremental costs' vs 'incremental benefits' and referred to as the Incremental Cost-Effectiveness Ratio (ICER). Within the economic evaluation process ICER are calculated per patient for all intervention options and allow for analysis and comparison of which intervention options provide the most benefit for the lowest direct cost. For example CEA can only be used to rank interventions in terms of cost per unit of effect along a single outcome dimension that is directly comparable between interventions. However, the outcomes of two interventions may not always be directly comparable. Although CEA is a straightforward method and can be especially useful when comparing outcomes that are not under question, its applications are limited. CEA cannot be used to compare multiple outcomes of an intervention, or when the outcomes differ in nature and therefore they cannot be compared directly (Brazier et al., 2007, Drummond et al., 2005). One solution to this comparison problem is cost benefit analysis (CBA).

2.2.2 Cost benefit analysis (CBA)

CBA is an approach in which all of the benefits of an intervention are measured in terms of monetary value, in theory it provides insight into the absolute benefit of treatment programmes. Outcomes such as increased survival rate or life years are valued using money as the numeraire. An intervention is considered worthwhile if the monetary valuation of all the benefits exceed the costs of providing the intervention. An advantage of CBA is that the measure of benefit incorporates a wider range of benefits and in particular non-health specific benefits. There are several techniques for calculating the monetary valuations of benefits. One technique is to impute values from people's 'revealed preferences' that are collected in a marketing setting. This is regarded as the most appropriate method to employ were feasible as actual decisions are assumed to be a more valid representation of peoples preferences, as opposed to asking the consumer what their preference would be. However, revealed preferences are difficult to collect in healthcare setting as consumers of healthcare services generally do not know the costs of the healthcare they are consuming, and often the consumers of healthcare do not pay directly for the healthcare that they receive in publicly funded systems like the UK NHS. The CBA method is based on the idea that those who gain from receiving the intervention will compensate those who lost out from funding the intervention. For example, if the taxation paid by a healthy individual provided the public funds to pay for the treatment of an unhealthy individual then the healthy individual may be compensated for this payment by the unhealthy individual returning to work and paying back into the social system. It is argued that by using money to measure the benefit gained, CBA can be used to address whether a treatment is worthwhile on a societal level. Using money to measure benefit also encompasses non-health-related benefits. However, asking people how much they personally would be willing to pay for another individuals' health can be challenging and uncomfortable for the respondent. (Drummond et al., 2005, Brazier et al., 2007).

2.2.3 Cost consequence analysis (CCA).

Another alternative approach is cost-consequence analysis (CCA). The results of CCA are presented in a table and all relevant factors are presented, not one single outcome. CCA does not rank the interventions, the ranking task is left to the decision maker. Therefore, this method produces helpful data for the decision maker and is closer to the traditional goal of economic evaluation of being an aid to decision making. CCA is not viewed as helpful as the scores on health status measures have no obvious intuitive meaning, CCA can be quite subjective and the basis for a decision can be unclear and often will not be based on explicit values. Score differences cannot be compared across dimensions and health status measures cannot be compared with other outcomes such as cost or survival rates. (Drummond et al., 2005, Brazier et al., 2007).

2.2.4 Cost Utility Analysis (CUA)

These techniques all have their advantages and disadvantages and the best method to use has been highly debated (Brazier et al., 2007, Bhattacharya et al., 2014, Drummond et al., 2005). However, the most widely used method of economic evaluation is cost utility analysis (CUA) (Brazier et al., 2007, Robinson, 1993). As with CEA, CUA compares interventions using the incremental cost per unit of effect measurement. In CUA the unit of effect is a QALY, which is a value that combines the duration and quality of life of an individual with a certain health state (Drummond et al., 2005, Brazier et al., 2007, Sassi, 2006, Bhattacharya et al., 2014). The QALY was developed in an attempt to combine changes to both the length and the quality of life and to provide a single number that can be used to compare the health benefits of treatment and technologies.

A QALY is calculated by multiplying an individual’s life expectancy by a value of health-related quality of life, or utility, that is measured on a scale from zero to one, with one being perfect health and zero being death. Here the term utility is used to refer to the preferences of individuals or society for a particular set of health outcomes; the more preferable the outcome the more utility associated with the outcome. This utility value can be negative as it can be viewed that there are health states worse than death. If being in the health state affects more than one dimension of health, such as physical effect and psychological effect, then the health-related quality of life value will reflect this. For example, those patients who are receiving hypertension medication may be assigned a utility value of 0.5. Spending ten years receiving hypertension medication can therefore be expressed as 5 QALYS; i.e. ten years in that health state is equivalent to spending five years in a state of perfect health. This QALY value includes the effect of hypertension medication on physical health, psychological health and the relative health state difference between requiring hypertension medication and full health. For more complex health profiles that involve a transition through different states over time, the overall number of QALYs will be calculated by combining the quality of life values of the different state (Drummond et al., 2005, Brazier et al., 2007, Bhattacharya et al., 2014). The QALYs can then be combined with the cost of the treatment to calculate the cost per QALY gained from the treatment. This cost per QALY value can be used to compare the cost-effectiveness of any treatment across all areas of health (Bhattacharya et al., 2014, Drummond et al., 2005, Brazier et al., 2007).

The advantages of using CUA to assess healthcare interventions lie in being able to assess more than one outcome and compare interventions with different outcomes that treat the same disease by using QALYs. As with CEA, it is possible to calculate an ICER using CUA when comparing two interventions which is alternatively referred to as the incremental cost per QALY. The incremental cost per QALY is the extra cost of an intervention over the next best alternative intervention available divided by the extra QALYs gained from the intervention (Brazier et al., 2007).

2.3 Study designs in economic evaluation

It is possible to measure the direct monetary costs of the treatment in terms of the resources that need to be consumed in order to provide the new treatment or technology. However, for health economists to calculate the health-related benefit of the treatment or technology in a general way so that QALYS can be calculated, it is necessary to first describe the health states relevant to the interventions and then measure the effects of the interventions being compared on the health state of an individual. The World Health Organisation (WHO) defines health as “a state of complete physical, social and mental well-being, and not merely the absence of disease or infirmity” (World Health Organisation). The WHO also describes health as a resource that is a vital component for the economic stability of both an individual and society. In terms of societal effects of providing or not providing an intervention the costs are harder to measure as these are often ill-defined concepts such as costs to the education sector for providing additional support to individuals with hearing loss, or a loss of productivity within the workforce due to unemployment or premature retirement. There are two popular methods of describing and measuring health states and their influence on an individual’s quality of life. When an individual is currently in the health state that is being measured, their health is often valued using health valuation questionnaires. If the individual providing the valuation of health is not in the health state of interest, the health state can be communicated using a textual description and then valued using a preference elicitation method.

2.3.1 Health valuation questionnaires

Questionnaires can be generic and cover an individual's overall health, or they can be disease specific and ask questions about specific domains of health affected by and relevant to a certain condition. Generic health questionnaires are used in multiple contexts: in a research setting, in clinical practice as an assessment tool, and for the purpose of carrying out economic evaluations of health states and interventions (McDowell, 2006 Keszei, Novak & Streiner 2010). Questionnaires normally comprise of multiple categories or domains of health. In order to systematically measure the impact of a health state on these different domains of health, such as pain or mobility, there will be multiple items measuring each domain (McDowell, 2006). The responses that are given to each item will have a numerical value applied to it and the domain will then be given a score based on the average of the items in the domain. Using the scores for the different domains of health being measured, an overall score for an individual can be calculated.

Generic measures of health are popular among researchers and organisations as the utility valuations of different health states calculated using data from these generic health questionnaires are directly comparable regardless of the specific disease being studied or the nature of the intervention(s) being evaluated. Theoretically driven questionnaires are more popular among researchers, as theoretical scale items can come from patients' opinions, observations of the target group, clinicians and research so are viewed as well validated (Keszei, Novak & Streiner 2010). Two theoretically driven measures of generic health that are particularly popular in both the fields of research and economic evaluation are the Health Utilities Index Mark 3 (HUI3) (Feeny et al 2002) and the EuroQoL EQ-5D health questionnaire (EuroQol Group 1990). Numerical utility values for health for the HUI3 are derived using the standard gamble methodology (see 2.4.1), and numerical utility values for health for the EQ-5D are derived using the time trade off methodology (see 2.4.2).

The EQ-5D is the preferred instrument of NICE, who make recommendations to the NHS and some private health care providers about Health Technology Assessment and economic evaluations (NICE, 2013). The EQ-5D domains of health are: mobility, self-care, usual activity, pain and anxiety/depression. All health valuation questionnaires are useful, but as with any method they also have their disadvantages. The major barriers to the widespread development of health-valuation questionnaires are mainly the response rates that questionnaires achieve (Asch et al., 1997, Baruch and Holtom, 2008, Baruch, 1999), the resources that are needed to carry out large scale questionnaire studies, along with the inability of researchers to access the reliability of sample especially if the sample has been collected using online methods (Wright, 2005).

In an attempt to combat this issue the EuroQol group, the group that created the EQ-5D, collected reference data from a representative sample of societies in various different countries. However, it has been argued that the reference data for the UK, calculated based on the preferences of a random sample of UK residents collected in 1994, may no longer be suitable for use as it is now quite dated and society and its views on health care and quality of life may have changed since that time (Kobelt, 2013, Mulhern et al., 2014). Therefore, researchers have questioned whether the reference data should be updated to improve its relevance to modern day assessments of health technologies and services (Pickard, 2015, Mulhern et al., 2014).

It may be that these questionnaires that measure general health are too general and do not tap into aspects of health affected by certain conditions. The EQ-5D has been criticised in the past for being insensitive to changes between cultures and countries (Norman et al., 2009). Also, as the EQ-5D is one of the smaller health valuation questionnaires in terms of the number of health dimensions, it has been viewed by some researchers as not being appropriate for assessing hearing (Grutters et al 2007), vison, and energy/tiredness, as the questionnaire has no items that focus on these areas of health (Brazier and Rowen, 2011, Longworth et al., 2014). The HUI3 is often used by those researching hearing as it is viewed as more sensitive to both hearing-related and health-related quality of life (Longworth et al., 2014). The HUI3 is a very flexible questionnaire with several different versions to cover self-administration, administration via interview, or administration via proxy, and administration of the questionnaire to adults or children. The HUI3 measures the current state of a respondent (McDowell, 2006). The HUI3 is intended for use in clinical outcome measurements, health surveys, programme planning and resource allocation. It produces values on a zero to one utility scale and can therefore be used in CUA to calculate QALYs. The HU3I classifies health states by recording functional capacity, its coverage is restricted to physical and emotional aspects of health and items in the questionnaire cover: vison, hearing, speech, ambulation, dexterity, emotion, cognition and pain. However, a limitation of applying the HUI3 to economic evaluations in the context of the UK is that the utility values it provides are based on the preferences of a sample of the Canadian public.

If the general health questionnaires are not flexible enough to tap into the aspects of health for a particular health state, then a disease-specific questionnaire may be used instead to measure the impact of a rarer or more complex health condition. A literature search on the condition of interest will most probably reveal several existing scales. If an existing scale doesn’t suit research needs, researchers will often create their own scales that they feel are more suited to addressing their research questions. A researcher would assess the appropriateness of existing scales by assessing the scales validity and reliability. A scale is deemed to have face and content validity for a specific research project if it assesses the desired aspects identified in the research question. Scale sensitivity and how small of a change it can measure will be of interest and a numerical measure of a scale's reliability can be gained. The scale must be stable: to be stable a scale needs inter-observer reliability (agreement between different observers) and intra-reliability (agreements made by the same person on two different occasions) and test-retest reliability. Scales also need evidence of construct reliability which confirms that constructs in the new questionnaires will be similar in scoring to other scales or to empirical findings (Streiner and Norman, 2008).

In summary, although questionnaire measures are advantageous due to their ease of application as they can be administered easily to multiple people or groups of people simultaneously (McDowell, 2006), their generalisability can also make them insensitive and not applicable to rarer or more complex health states. Another approach is therefore sometimes needed to describe and measure health and its effects on individuals.

2.3.2 The evolution of health valuation questionnaires

Numerous changes have been made to the commonly used health valuation questionnaires, the HUI3 and the EQ-5D. With the HUI going through three different versions to reach its current iteration (Horsman et al., 2003), and the original Euro-Qol instrument had evolved into a shorter three level instrument, the EQ-3D-3L, and longer five level version developed, the EQ-5D-5L. There has also been the generation and development of various different language versions (Devlin & Brooks 2017). Recently there has also been research investigating the possibility of additional supplementary aspects that could be added to the EQ-5D (Yang et al 2014, Yang et al 2015, Devlin & Brooks 2017). These supplementary items have been called 'Bolt ons', and a 'bolt on' is an additional dimension added to the EQ-5D which describes specific health problems. Researchers have attempted to create bolt ons for a variety of health conditions and dimensions including: sleep (Yang et al 2014), vison, hearing and tiredness (Yang et al 2015). The success of these bolt ons has been varied; the addition of a sleep bolt on to the EQ-5D was found to have no significant effect on utility scores (Yang et al 2014), bolt ons for vison, hearing and a tiredness had varying impacts on utility values, dependent on the level of difficulty these conditions presented, i.e. moderate problems or severe problems (Yang et al 2015).These bolt ons were tested using the EQ-5D-3L, the varied success of the bolt ons when communicating difficulties with vision, hearing and tiredness could have stemmed from a lack of sensitivity that the EQ-5D-3L has, it simply may not be possible to fully communicate the impact of having a hearing loss, and of different levels of hearing loss and hearing conditions, using terms such as “I have no problems hearing, I have some problems hearing, I have extreme problems hearing”.

2.3.3 Text-based descriptions of Health

To overcome the issues of using generic or disease specific health valuation questionnaires, text-based descriptions (or ‘vignettes’ as they are also known) can be used to describe health states, following the presentation of a text description utility values can be obtained using specific methods for obtaining utility values such as standard gamble and time trade off (see 2.4 for further explanation). Text descriptions of health states make it possible to describe any health state to a member of the public who may have little to no knowledge or experience of the condition being described. For rarer or more unusual health states, researchers often choose to develop disease specific text-based descriptions of a health state where existing questionnaires may be inappropriate or insensitive.

A text-based description will usually detail the key attributes of the health state being described. Some researchers will not limit the attributes of the health states covered by the descriptions, leading to lengthy descriptions of health states (Lee et al., 2000). Other researchers will limit the attributes covered by their description from the outset in order to produce a description of a heath state that covers what they believe to be the essential information (Cadman and Goldsmith, 1986).

Text-based descriptions of health states can be developed in many different ways using different information sources to build the descriptions; e.g. patients, clinicians, the literature. A search of the literature looking at vignette creation across different areas of health demonstrates just how diverse the approach to creating vignettes can be. Table 1 shows some examples of studies that have developed vignettes. Researchers have used previous literature, their own experiences within research or clinical settings, clinicians’ views, patient views, family and friends’ opinions, guidelines, or a combination of any of those factors. The vignettes can also follow many different structural formats; they can be notes, bullet points, prose, descriptions from a third person perspective, separated into symptom type or separated into symptom severity. This wide-ranging approach to constructing a vignette to describe a health state leads to varied descriptions that have an equally varied rate of success in describing the target health state. This is why many of the studies detailed in Table 1 also went through a process of validating the health state descriptions, again using different methods, different information sources, and to different degrees of comprehension.

Study	Information used to create vignettes	Symptoms covered in the vignettes?
Lenert et al. (2004)	Clinical expertise of research team and other experts.	Primary symptoms
Vellinga et al. (2004)	Consultation with experienced geriatricians & psychiatrists	Primary symptoms.
Banks et al. (2014)	NICE guidelines, qualitative interviews with patients & clinical experience of a member of the research team. Vignettes were organised by symptoms by risk level (1%, 2%, 5%, 10%)	Primary symptoms.
Dusch et al (2014).	Survey of patient's surgeon preferences	Primary & Secondary symptoms (related to level of care by surgeon)
Rowen et al. (2012)	Based upon the European Organisation for Research and Treatment of Cancer core quality of life questionnaire.	Primary & Secondary symptoms.
Gerhards et al. (2011)	Guided by Diagnostic Statistical Manual-IV content.	Primary symptoms
Shingler et al (2013)	Review of published literature, exploratory interviews (with clinicians and patients), draft health states were reviewed by clinicians, patients and members of the public.	Primary & Secondary symptoms.
Osborne et al. (2012)	Initial vignettes were designed by experienced researcher and based upon literature. The vignettes were then reviewed by members of research team and then reviewed by psychiatrists, members of a peak consumer organisation, in this case the charity SANE, and patients.	Primary, Secondary & Tertiary symptoms.
Stein et al. (2007)	Based upon points used in validated quality of life measure for ovarian cancer.	Primary symptoms.
Doyle et al. (2008)	Literature review to form draft health states. Interviews with lung cancer experts to validate health states. Interview results analysed by two independent researchers & two psychometric experts. Health states piloted with members of general public in Standard Gamble interview.	Primary & Secondary symptoms

Nafees et al. (2008)	Literature review. Interviews with experts. Health state first drafts reviewed by original experts and new experts & two psychometric experts. Piloted with members of general public.	Primary & Secondary symptoms.
Lloyd et al. (2006)	Literature review, cycles of exploratory interviews, focus group with expert physicians & further content validation interviews. Health states piloted with members of general public in Standard Gamble interview.	Primary, Secondary & Tertiary symptoms.

Table 1: Studies that developed vignettes to describe various health states.

2.4 Methods of obtaining utility values

Once a text description of a health state has been developed and refined the researchers can then ask members of the public to make preference judgements about the extent to which they would prefer to live in perfect health rather than to have the quality of life of someone living in the described health state. There are three methodologies used with text-based descriptions and health valuation questionnaire methods to obtain the general public's preferences for certain health states or conditions relative to perfect health.

2.4.1 Standard Gamble (SG)

The SG technique is considered the classical method for collecting utility values and was first developed by von Neumann and Morgestern in 1953 (as cited by Torrance, 1986). With the SG technique, the person rating the health states is offered two options: the certainty of staying in the described health state for the remainder of their life or the gamble of a treatment option. The treatment option will have a probability of delivering either perfect health or death. The probability that the treatment will result in death is altered throughout the task and when the probability, or the gamble, that the respondent would be taking when they chose between death and perfect health is seen as indistinguishable from living in the health state then that probability value is the estimated quality weight of the described health state (Torrance, 1986, Bhattacharya et al., 2014). The main issue with using SG as a utility collection method is the technique's use of probabilities. Some reaserchers have used visual aids to combat these difficulties which have included probability trees, probability wheels, tables, and figures that show faces to represent the probabilities of death (Torrance, 1986, Clarke et al., 1997, Bhattacharya et al., 2014). Even with the inclusion of visual aids to assist with the portrayal of probabilities and the gamble being offered, some reseachers think that that the SG method is not an approachable method to use with the general public when investigating health states and utility values (Torrance, 1986).

2.4.2 Time Trade Off (TTO)

TTO is another preference-based method of obtaining utility values. It is considered to be a less complex alternative to SG as TTO doesn't use probabilities to obtain preference values (Brazier et al., 2007, Torrance et al., 1972, Torrance, 1976, Torrance, 1986). The TTO approach presents a hypothetical situation to the individual rating the health state. They are informed that they will live in the health state described for a certain number of years and then they will die a painless immediate death after that time. The person rating the health state is then given the choice of giving up years off the end of their life to return to a state of perfect health for the remaining years. Usually when using the TTO approach, the time period specified for trading is ten years left to live in the described health state. However, there is no general consensus on how long this trade off (or 'tariff') period should be, with time periods used in the literature varying from one month to several years (Attema et al., 2013, Brazier et al., 2007, Stalmeier et al., 2001). Attema et al. (2013) found that although there is great variation in the tariff length used in TTO studies, most used the trading time period of ten or twenty years.

The trading tariff length, and the way the trading question is posed to the person doing the trading can significantly influence the trading values obtained (Arnesen and Trommald, 2005, van Nooten and Caro, 2013, Attema et al., 2013).

Several hearing-related studies have utilised the TTO method with a ten year trade off tariff (Abrams et al., 2002 as cited by Hosford-Dunn, 2011; Summerfield et al., 2002). For example, when using the ten-year TTO approach in hearing research a participant may be given a text description of profound deafness, and they will then be asked to imagine that the description details their own health state. They will then be asked to imagine that they have ten years left to live, living with profound deafness, the previously described health state. Next the participant is told that they have the option to give up some of these ten years in order to live the remaining years in perfect health. The greater the number of years they are willing to give up in order to not live in the health state then the worse the quality of life the health state is viewed to be and the lower the health utility value is for that health state. If the respondent is willing to give up four years of their hypothetical ten years of remaining life, then they would live six years in perfect health before dying. To calculate the health utility value in this example, where utility is just another way of saying preference, the following formula would be used:

$$\frac{10 - x}{10}$$

where x refers to the number of years that the respondent traded in order to live in perfect health. In this example, the individual was willing to trade four of their ten years to live in perfect health and therefore the utility for the described health state was judged to be 0.6.

2.4.3 Visual Analogue Scales (VAS)

Unlike SG and TTO, a VAS scale (sometimes called a category rating scale or simply a rating scale) is not a preference-based method of collecting health utility values. A VAS can use a range of values to collect health ratings; e.g. zero to one, one to ten or one to one hundred. The end points on the scale can also be defined as many health levels. For instance, the lowest point on the scale could be defined as representing death, poorest health imaginable or poorest quality of life and so on. Equally, the highest value on a VAS scale can refer to excellent health, perfect health, or good quality of life and so on. The span of a VAS scale and the endpoint definitions vary between researchers, research group, and research areas (Brazier et al., 2007).

When a VAS is used to gain health utility values from a text description of a health state, participants are asked to indicate on the scale how good or bad the participant thinks their health is or the health of the person described in the health state is. VAS values collected in this way are assumed to be interval data, so if the scale runs from zero to one hundred then it is assumed that the difference between ten and twenty is the same as the difference between thirty and forty (Brazier et al., 2007). As discussed by Summerfield et al. (2002), if VAS scales are being used to collect health utility values in a economic evaluation then the values collected will have to be converted for a researcher to be able to use the values to represent health utilities using the following formula:

$$U_{VAS} = 1 - (1 - V/100)^{1.6}$$

where the exponent of 1.6 can be varied to map onto utility data obtained using TTO or SG methods. In summary, adjusted VAS values can be used to obtain health utilities where concerns exist about the ability of respondents to complete more complicated preference-based methods, but VAS is typically used alongside other preference-based health utility methods, namely TTO and SG.

2.4.4 Methodologies used in research

There has been a lot of interest within the literature in evaluating the utility collection methods and comparing the three methods to determine the best utility preference measure to use when evaluating a health state. Read et al. (1984) compared VAS, TTO and SG collection methods for coronary related health outcomes. They found that utility values collected using all three methods had a certain level of consistency, but that there was not overall consistency in utility values collected using VAS SG and TTO methods. Dolan et al. (1996) concluded that TTO and SG are efficient and valid methods of collecting utility values, but that it may be more beneficial to use TTO with visual aids, as this collects more complete valuations than SG both with and without visual aids. Overall, it appears that all three methods are valid, reliable, and consistent methods to use when collecting utility values. However, the general public, who are often the people called upon to rate health states, have preferences over the method used. Lenert and Sturley (2001) looked at VAS, TTO and SG methods and acceptance of the methodology. They found that individuals prefer to use VAS and TTO methods over SG and they recommended that VAS and TTO be used together in valuation studies. The preference for using TTO and VAS could be down to the use of probabilities in the SG method and a difficulty in understanding the gamble being offered, as previously discussed. This does suggest that a combination of VAS and TTO may be optimal when carrying out economic evaluations of health technologies and healthcare services as they are generally better understood by respondents than SG.

2.5 External factors that could effect health state valuations

2.5.1 Focussing illusion

Ubel et al (2001) explored whether focussing illusion could cause non-patients to overestimate the impact of a health condition. Focusing illusion occurs when people are asked to imagine themselves in unfamiliar circumstances and they focus on what would be different, overlooking what would stay the same. By attending asymmetrically to factors that would produce changes in well-being people can overestimate the significance of those changes.

There are reported discrepancies between patient obtained valuations of health states and non-patient obtained valuations, for example Ubel et al 2001 cited a study by Brickman, Coates & Janoff-Bulman (1978), in which individuals with paraplegia stated that the condition has a relatively small effect on their lives and that the happiness of individuals a year after developing paraplegia was nearly the same as the happiness of people a year after winning the lottery.

2.5.2 The influence of lived experience

There is evidence within the literature that when individuals live through and experience a health state, this alters their view of the health state and its effects on their QoL. This is often referred to as response shift. Jansen et al. (2000) investigated the quality of life of breast cancer patients before and after radiotherapy. They found effects of living through the health state with a ‘scale recalibration’ observed within the patients’ own internal standard that would have significantly altered the quality of life evaluations they gave themselves pre-treatment to post-treatment. Prior to experiencing treatment, the impact of radiotherapy on fatigue and overall quality of life would have been underestimated by patients.

Clayton and MacKay (2017) highlighted that there is often a difference between decision utility metrics such as the HUI 3 and EQ-5D, and experienced utility metrics. They argued that these decision utility metrics should be employed in conjunction with experienced utility metrics and that respondents to decision utility surveys should be provided results from experienced utility metrics measuring the same disease. Clayton et al (2017) argued that providing the experienced based utility metrics would result in respondents making better informed decisions and improve QALY calculations. Thus enhancing the decision-making of policy makers when allocating scant resources to support the populations’ health needs. This stance does pose an interesting question, that researchers could go one step further and could combine the decision based and experience based utility metrics by using simualtions’ to provide respondents with near direct experience of the target health state?

2.6 Hearing loss

Hearing loss is measured in terms of the type of hearing loss: conductive or sensorineural hearing loss, and the severity of the hearing loss: ranging from mild to moderate to severe. A conductive hearing loss occurs when sounds can no longer reach the inner ear, this means that somewhere within either the ear canal, the ear drum, the middle ear and the ossicular chain, obstacles have occurred that prevent sound from reaching the inner ear. To combat a conductive hearing loss sound must be made louder to achieve a normal stimulation of the inner ear. A sensorineural hearing loss is caused by changes in the inner ear and the first part of the hearing nerve. Most commonly this is due to damage to the hair cells found within the inner ear. Inner hair cells transmit sound information to the hearing nerve, these cells connect to around 28,000 nerve fibres, the outer hair cells connect to about 1,500 nerve fibres. When the outer hair cells are set in motion by sound vibrations, they contract and expand. These movements of the outer hair cells amplify the stimulation of the inner hair cells (Elberling & Worsen 2006).

2.6.1 Unilateral hearing loss

Unilateral hearing loss, or Single Sided Deafness (SSD) is defined as normal hearing in one ear and impaired hearing in the other ear and can range from a mild to severe loss. The hearing loss can be conductive, sensorineural, or a combination of these factors (Elberling & Worsen 2006). In the literature an interaural pure tone average (frequencies 0.5, 1, 2, and 4 kHz) difference of more than 10 to 15 dB HL has been used to define parameters for SSD (Noble & Gatehouse, 2004). In an international consensus statement Vincent et al, (2015) defined severe SSD has as a pure tone average ≥60 dB hearing loss in one ear.

Prevalence rates for SSD have been estimated to be around 1% of the general population (Davis and Research, 1995) SSD patients have difficulties with sound localization and listening to target sounds in background noise (Douglas et al., 2007). Individuals with SSD experience a clinically significant degree of audiological disability due to the impaired function of one ear (Chiossoine-Kerdel et al., 2000). The impact of SSD on an individual's QoL has been investigated within the literature, however, often the impact is assessed using disease specific measures or bilateral deafness is assessed. Arndt et al. (2011) asked unilaterally deaf adults to complete the HUI3 prior to receiving a CI, the average utility score for this pre-implantation was 0.56. Summerfield et al (2002) collected utility values using the HUI 2 from individuals prior to their unilateral cochlear implantation, the average utility value for this group was 0.75.

There are possible treatment options for SSD depending on the level of hearing loss present, these include hearing aids, contralateral routing of signals (CROS) aids, bone anchored hearing aids (BAHA) and cochlear implants (CI). A hearing aid is a device which amplifies sound and directs the sound through the ear canal to the inner ear. A CROS aid is device which reroutes sound from the side of the impaired ear to the hearing ear. A BAHA is fixed to an implant in the skull and conducts sound through the skull to the inner ear, and a CI is a surgically implanted prosthesis used to restore some level of hearing to patients with hearing loss (House 1976, Clarke 2004, Elberling & Worsen 2006, Schrøder, Ravn, & Bonding 2010, Lucas et al 2018). SSD is a main research interest of the Cochlear Implant research group within the Nottingham BRC.

2.6.2 Text-based description of hearing loss

There have been several studies in recent years that have used text-based descriptions to describe hearing-related health states. One such study is Summerfield et al. (2002) in which four vignettes were used in order to estimate utility values for four health states related to bilateral Cochlear Implantation in adults with profound deafness. These vignettes described hearing-related health information such as speech perception levels and psychosocial functioning. As with many studies that use vignettes, the researchers did not explain where the descriptions came from, nor do they detail any aspects of the development of the vignette. However, the health valuations obtained in this study were later found to be similar to those elicited directly from CI users, suggesting that the text-based descriptions were successful in communicating important features of the health states (Bond et al., 2009). In a more recent study, Summerfield et al.

(2010) again used vignettes to investigate and assess the cost effectiveness of Cochlear Implantation in children with profound deafness. This study used previous studies, experience, research reports, reported issues from parents and online discussion boards to develop descriptions of hearing-related health states. The vignettes described the everyday functioning and prospects of a child with bilateral profound deafness.

Lucas et al (2018) established a text-based description of the impact of living with SSD, the description was developed by consulting with eight individuals with a clinical diagnosis of SSD, who had had the diagnosis for longer than a period of 12 months and whose hearing level had been stable for the last 12 months. Lucas et al (2018) developed a validated description of living with SSD and within this process they reported a range of themes which describe the different aspects that individuals with SSD report having difficulty with (functional consequences of SSD, psychological consequences, social consequences and the positive consequences), and example situations in which individuals with SSD said they struggled with.

2.6.3 Methods of obtaining utility values used in hearing research

In scenario-based hearing research, there have been several studies and all have taken a different approach to methodology usage. Cheng et al. (2000) used a combination of VAS, TTO and the HUI3 to determine QALY values and cost consequences of cochlear implantation in deaf children. No vignettes were used in this study. Instead, the TTO task given to the parents of deaf children was a choice between the child remaining in their current health state or trading years of life to live in perfect health.

Summerfield et al. (2002) has also used a health valuation questionnaire, namely the HUI Mark II, but chose to only pair that questionnaire with a TTO task. This study aimed to perform a cost-utility scenario analysis of bilateral cochlear implantation in adults. Summerfield et al. (2010) used TTO and VAS to estimate the increase in quality of life gained by deaf children from bilateral cochlear implantation compared to the increase in quality of life gained from implantation in one ear only. The study also aimed to estimate the additional cost to the healthcare system in the United Kingdom for providing bilateral implantation. It is worth noting that Summerfield found that it was difficult for individuals with hearing loss to evaluate hearing-related health states. This is not because the methods used were complex but because individuals with hearing loss do not always see themselves as ill or not in perfect health. This could be an issue hearing researchers encounter when using the TTO methodology in the economic evaluation of hearing-related health.

Bokari et al. (2010) used a VAS as a method of patients self-reporting hearing loss with some degree of success. In addition to this particular study, VAS has also been used in studies that have evaluated health utilities in hearing and has often been used in conjunction with other preference based health utility methods and with health utility questionnaires (Lovett et al., 2010, Kuthubutheen et al., 2015).

2.7 Potential use for multimedia in the health valuation process

In section 2.3.2 the use of text descriptions, or 'vignettes', to describe health states so that they may be valued was discussed. However, there has previously been an interest in the use of multimedia techniques in the health valuation process, particularly if the use of multimedia in presenting health state descriptions. One research group in particular has conducted extensive research on the possibility of utilising an emerging technology, multimedia, in the economic evaluation process.

Examples of the use of multimedia to describe health states are apparent as early as the 1990s with Goldstein et al. (1994). This particular study looked at the possibility of using multimedia in the presentation of health state descriptions to be assessed by healthy individuals, not patients. The study aimed to investigate the impact of the medium used to present a health state on understanding of the health state and on utility values. Goldstein and colleagues used a computer programme to present multimedia presentations (audio clips and photographs in this case) of a health state relating to Gaucher disease, along with a text only representation to allow for a comparison between the two mediums. The researchers hypothesised that using multimedia to present a health state description would: (a) focus a respondent's attention onto the content of the description and remove reading problems; (b) lead to greater recall of information presented; (c) lead to greater recognition of information presented in the health state descriptions; and (d) result in smaller 'preference intervals' in the VAS ratings collected. In this study the researchers referred to recognition as the ability to correctly identify the information that had been presented to them in the description of the health state. Preference intervals on a VAS was a method developed by the researchers in order to measure the uncertainty in individuals' VAS responses. The rationale behind preference intervals was that a participant would give two ratings on a VAS scale, a point on the scale that they believed to represent the lowest quality of life for the individual in the description and a point on the scale that they believed to represent their highest quality of life. The health state described was of Gaucher disease, which is an inherited enzyme deficiency disorder which causes an accumulation of excessive glycolipids in the individual's spleen, liver and bone marrow. The description was written in first person, as the researchers believed that this would lead to the described health state being viewed as more representative of a real life individual and not a hypothetical scenario. The vignette that Goldstein and colleagues created covered physical health, social functioning and role functioning. It did not cover emotional health or self-rated health; the researchers did not justify this decision but did say that participants were left to make their own inferences on the individual's emotional health and their views of their own health. The vignette was developed using medical literature and was worded as if it were coming from a child.

The multimedia used to present the health state was photographs of a child presented alongside audio of a child reading out the health state descriptions. The photographs changed to other relevant photographs as the audio of the description continued. The multimedia material was one minute long in total and presentation of the multimedia was repeated once.

Participants then completed a free recall task of description, a multiple-choice recognition task and a preference interval rating task. Free recall was measured as how many of twenty-four conceptually distinct segments the participant recalled. Recognition memory was measured using multiple choice questions pertaining to the descriptions. Goldstein and colleagues concluded that the multimedia group recalled more of the descriptions presented using multimedia, and that there was a higher recognition of information presented in the multimedia condition. The researchers also noted that participants inferred things from the photographs that were not stated in the text descriptions, i.e. that the boy appears to need oxygen, or that the boy appears to need a feeding tube, or that the boy can't be ill as he smiles. Both groups, the text-based health state presentation condition, and the multimedia-based health state presentation conditions, drew inferences on the health of the child based on what was presented to them. The researchers also concluded that the quality of life values obtained did not differ based on the medium used to present the health state, although this was not a statistically significant result. As the research team concluded that the quality of life preference intervals were not statistically significant, the researchers have appeared to focus mainly at the recall and adequacy of the health state descriptions in their discussion of the study results. A potential shortcoming of this study was that the health state description was not informed by individuals with the condition described, medical professionals or experts nor was the final description verified by any consultation. It is possible that the particular choice of imagery used had a strong influence on the health valuations. Specifically, the use of a child to illustrate the health states may have created bias. The choice of multimedia material and vignette tone (being developed as if it were coming from a child, using a male child in the multimedia, emotive photographs) could have influenced the quality of life scores obtained from the VAS scale, the rare nature of the disease could also have affected the participants' understanding and comprehension of the description.

Morrs et al. (1994; cited in Goldstein et al., 1994) developed a patient educational tool that also measured health utility values of anti-psychosis medication in schizophrenia patients. Morrs and colleagues took a different approach to that of Goldstein et al (1994) by using short video sequences accompanied by audio recordings to describe three different possible side effects the patients could experience. The patients' preferences for not living with these side effects were assessed using VAS and SG. Instructions were presented to the patients using both voice recordings and in print, and, if necessary, were clarified by a moderator. This approach was adopted to ensure that the patients, a vulnerable participant group, fully understood the side effect being rated. Animated displays were used to graphically display probabilities in the SG task using a wall of faces. A box with a hundred faces was used and for each percentage increase in the probability of death, one face would be replaced at random by a cross. The researchers also used the more traditional pie chart method of displaying probabilities. Participants were also given comprehension tests and validation questions throughout the experiment to verify comprehension and attention. The development of the descriptions of the three different side effects that are used in this study were developed from medical literature. Although Goldstein et al. (1994) did not look at the use of multimedia as an educational tool as Morrs et al (1994) did, Goldstein et al (1994) did expand upon Morrs and colleagues research by looking to use multimedia in health utility valuations in health populations and by comparing multimedia usage to traditional text description methods.

There has also been research to develop computer programmes that can be used to collect utility values, with U-titer, developed by Sumner at al 1991, ADLIB, developed by Goldstein et al. (1993) and more recent programmes like impact4 developed by Bansod et al (2003). These programmes aim to further digitise and streamline the utility collection process, which has been continued in recent research using technology to collect TTO data (Attema et al., 2013). Clarke et al. (1997) used similar methods to obtain utility values as Goldstein et al (1994) to test the effects of a search procedure on utility scores. Search procedures in this context refer to the method used to determine the point at which a respondent is indifferent between the health state described and the trade or gamble being offered. Clarke et al. (1997) used specifically built software when carrying out this study, supporting the use of computer programmes to collect utility values.

However, the use of multimedia is not without challenges. Lenert et al (2000) assessed the risks of using multimedia methods in educational tools and decision support. The research portrayed the same health state but used actors of different racial backgrounds and different genders to portray the heath states. The study found that there was an effect of actor race on utility values collected using VAS, with utility ratings found to be lower when the race of respondent and actor differed. When using the SG method, an effect of gender was also found with female actors producing higher utility values compared to male actors. As the research was conducted within the US, it is unclear whether similar effects would arise in other cultures such as the United Kingdom where societal views on race and gender may differ from those of the respondents in that study. The observed racial and gender related biases may also reflect the period when the research was conducted and may not now be replicated due to changes in society and its views over time.

These studies illustrate that multimedia representations of health states have been used to describe health states and to collect utility values. This use of multimedia is encouraging as it hints at the feasibility of continuing to use multimedia, such as visual or audio presentations of health states, in the economic evaluation process. However, there has been a lack of validation of multimedia methods in health state presentations. To date, there have been no studies that have looked at the possibility of using multimedia to present hearing-related health states. It is also worth noting that the technology used in these studies is now heavily outdated, and it could be that current technology, especially with the emergence of virtual reality and advance computer programming software, would allow for more sophisticated multimedia presentations to be developed.

2.8 Rationale for the thesis

Multimedia utilisation opens possibilities for using new technologies to provide more comprehensive descriptions of health states to respondents who may have no experience or knowledge of the heath state being described to them. The current research project aims to explore alternative methods of describing the heath states related to hearing that have been prompted by recent advances in health valuation and multimedia use, namely presenting additional forms of information in novel modalities such as videos of patient interviews, to provide a more comprehensive description of a health state to the informants who are tasked with valuing it (Goldstein et al., 1994, Lenert and Sturley, 2001).

What is relatively unique about hearing loss compared to many medical conditions is that it is possible to simulate many of its effects in normal-hearing individuals through the use of computer-generated 'acoustic stimulations' of hearing (ASH). The effects of interventions such as hearing aids and CI could also be simulated acoustically. Acoustically simulating hearing loss is now feasible with the improvements in technology that have occurred since Lenert and colleagues looked into multimedia use in health state valuation across the 1990s and early 2000s (Goldstein et al., 1994, Goldstein et al., 1993, Lenert et al., 1995, Lenert and Soetikno, 1997, Lenert et al., 2000, Lenert and Sturley, 2002)

There are many study designs in economic evaluation that could be used to investigate the potential use of ASH to momentarily describe the effects of listening with a hearing loss and to aid the description of hearing-related health states within health valuation exercises. Health valuation questionnaires are already widely used within the economic evaluation of health and are well validated. Some research has investigated the addition of extra items or 'bolt-ons' to the EQ-5D (Yang et al 2015), which suggests that using simulations of hearing in a way similar to the additional questions would not be implausible. However, this method may require several different ASH to be developed to describe relatively abstractly described levels of hearing difficulty. The patient validated text-based description developed by Lucas et al (2018) lists within the description specific listening situations which individuals with SSD find difficult, using this description to inform the development of some initial ASH would ensure that the ASH developed were relevant and realistic of the listening problems experienced by individuals with SSD. Describing the impact of a hearing-related health condition using simulations could facilitate accurate QoL valuations which are in line with patients' self-reported QoL, without relying on the conceptualisation abilities of the rating public.

3 Feasibility of using acoustic simulations to describe health states

3.1 Introduction

As previously discussed in Chapter 2 (Section 2.6.2), it is difficult to describe the impact of having a hearing loss on quality of life, either using a set of general health dimensions or with words alone. Therefore, it is not clear whether valuations of health states associated with hearing loss that are either described using health dimensions in questionnaires, or by text descriptions, could ever adequately reflect the impact of hearing loss on an individual's health-related quality of life. Previous studies have developed computer based utility collection tools (Sumner et al., 1991, Goldstein et al., 1993, Bansod et al., 2003) and have utilised multimedia representations of health states to describe health and to collect utility values (Goldstein et al., 1994). Simulating a hearing loss acoustically (i.e. through the presentation of auditory demonstrations) is now feasible with the improvements in technology that have occurred since Lenert and colleagues investigated the potential use of multimedia in the health state valuation process (Goldstein et al., 1994, Goldstein et al., 1993, Lenert et al., 1995, Lenert and Soetikno, 1997, Lenert and Kaplan, 2000, Lenert and Sturley, 2002). Presently, there has been no research conducted into the possibility of acoustically simulating hearing-related health states for the purpose of obtaining utility values from members of the public using preference-based methods. Using multimedia representations, or computer software, to build a computer-based simulation of a hearing-related health state would be feasible, a notion supported by the success of previous research in implementing multimedia tools to portray health states and given previous research using simulation to develop and improve hearing aids algorithms (Vicen-Bueno et al., 2007, Florentine et al., 1988). Therefore, it is reasonable to assume that simulating a hearing-related health state and different levels of difficulty with listening is an attainable goal and that using these simulations to aid with the health state description process is not implausible.

The overall aim of the work detailed in this chapter was to create and develop acoustic simulations of hearing (ASH); of a specific form of hearing loss, single-sided deafness (SSD), and to explore how feasible such ASH are for describing a health state when collecting health utility preferences for hearing-related health states. The objectives were to:

1. Develop ASH of two different listening situations that depict different levels of listening difficulty experienced by individuals with SSD
2. Investigate, through pilot testing, whether these ASH communicated an impact of hearing loss on health-related quality of life, and whether they were capable of communicating different levels of both difficulty and impact on health-related quality of life
3. Determine the how feasible the ASH are to use when describing a hearing loss in order to collect health utility data
4. Evaluate the effectiveness of the ASH in recreating the difficulties that an individual with SSD would experience

3.2 Initial Development of acoustic simulations

3.2.1 Hearing loss described and simulated

Hearing loss can be classified in terms of the type of hearing loss (conductive or sensorineural hearing loss) and the severity of the hearing loss which ranges from mild to profound (Elberling and Worsoe, 2006). Profound single-sided deafness (SSD) was chosen as the hearing-related health state to simulate in these pilot experiments. Profound SSD in the context of this experiment is defined as normal hearing in one ear (<30 dB Pure Tone Average Hearing Level) and profound deafness in the other ear; i.e. no access to acoustic information, this definition is in line with that provided by a consensus paper (Vincent et al., 2015). The parameters for normal hearing were selected so that the target hearing level (<30 dB Pure Tone Average) would reflect the typical range of hearing in individuals without a hearing loss in the general population, and this normal hearing level was used to define the required hearing levels of participants in the experiments. Profound SSD was chosen for two reasons. Firstly, profound SSD is a relatively straightforward form of hearing loss to simulate acoustically as it simply involves presenting sounds to one ear that has been confirmed as audiometrically ‘normal’ and presenting no sounds to the ‘deaf’ ear and that the initial simulations could be developed quickly with minimal technical investment. This approach also meant that the simulation of less profound degrees of hearing loss (where some acoustic information is received through an impaired ear) was not necessary. Simulating less profound degrees of SSD would have meant that simulations would need to be sufficiently complex in order to recreate the characteristics of a compromised auditory system, which would have been a lengthy and intricate process, and one that would have been difficult to validate.

Text descriptions, sometimes referred to as ‘vignettes’, represent a common method of describing a specific health state so that preferences for living in that state can be elicited from members of the public. An evaluation of the feasibility of acoustic simulations to describe health would therefore involve a comparison between preferences obtained using acoustic simulations to those obtained using text descriptions. A second reason for simulating profound SSD was therefore that the severe-to-profound hearing loss (SPHL) research group at Nottingham Biomedical Research Centre (Nottingham BRC) has worked extensively with SSD patients and has developed patient-informed and patient-validated text descriptions of SSD-related health states, both aided and unaided. Lucas and Kitterick (2014, Lucas et al., 2018) conducted focus groups with 8 SSD patients, in which the patients identified key situations that they have difficulty with. The data were then used to develop text descriptions of living with SSD and these were then validated by consulting with a larger group of SSD patients (see Appendix 1, section 9.1 for the full unedited descriptions). The use of these descriptions therefore allowed the current study to focus exclusively on the development and validation of acoustic simulations without having to also develop and validate new text-based descriptions. Furthermore, the views of the SSD patients collected through the focus group sessions performed by Lucas & Kitterick (2014) could be used to inform the acoustic simulation development process, such as the situations simulated and what these situations entail. The availability of these data meant that the initial acoustic simulations of SSD described in this chapter were informed by both the published literature and by patients' lived experience. Further qualitative work to inform the choice and content of acoustic simulations to describe this form of hearing loss is described in Chapter 7.

3.2.2 Situations simulated, set up and simulation environment

Research has often highlighted the listening difficulties faced by individuals with SSD, with Giolas and Wark (1967) outlining in detail the problems faced by individuals with SSD: difficulty understanding speech in quiet environments, difficulty understanding speech in noisy environments, and difficulty determining the location of, and the direction of sounds (localisation). Interestingly, the difficulties of modern day SSD patients could still be placed in these categories, despite improvements in care for SSD patients with more recent research identifying those patients reporting the same problems. Lucas & Kitterick (2014, 2018) found that patients continued to identify difficulty understanding, and following a conversation when more than one person talks at the same time even when there was no other sound present in the environment as well as problems listening to speech or following a conversation in a noisy setting. They also reported that patients had greater difficulty understanding, and following, a conversation when there was general background noise present. As discussed in Chapter 2 (Section 2.7), the difficulty that SSD patients have with listening to one person talk when other people are talking at the same time stems from an inability to determine the location of the speakers (localisation) without binaural hearing, hearing in both ears (Lucas and Kitterick, 2014, Arndt et al., 2011, Vermeire and Van de Heyning, 2009). Binaural hearing allows an individual's auditory system to determine the direction of the source of sound by using the differences in interaural level of the sounds and the time the sounds present at each ear (Akeroyd, 2006).

The main listening difficulties that appear to be reported consistently by individuals with SSD are with localising sounds and understanding speech (McLeod et al., 2008, Sano et al., 2013, Subramaniam et al., 2005, Wie et al., 2010, Lucas et al., 2018). Therefore, the two situations were chosen to simulate for initial pilot testing were: speech in speech with background noise (SiS+B) and speech in speech without background noise (SiS). SiS+B was chosen as a situation to simulate as it appeared to have a substantial negative impact on the patients' hearing function. SiS was also simulated as it had a lesser impact but still imposed some level of listening difficultly. These two situations would therefore cover a range of everyday problems that individuals with SSD experience and also provide a comparison that could be used to assess the effectiveness of AS at communicating different levels and types of listening difficulty.

The configuration of the acoustic environment to be simulated was also informed by the situations in which SSD patients experienced difficulty. Virtual acoustics were used to simulate a group conversation occurring in a cafeteria situation. A database of binaural room impulse responses developed by Kayser et al. (2009) was used to create the acoustic simulations. The binaural impulse responses were measured with an in-ear microphone at both ears of a human head and torso simulator. The database contains different sets of binaural room impulse responses for multiple realistic head and sound-source positions in four natural environments. A binaural impulse response is the response at the two ears to an impulse or 'click', the recorded pattern of sound at the ears reflects the acoustic properties of the room and encodes information about surface reflectivity and echo strength and timing (Kayser et al., 2009, Jeub et al., 2009). The available environments provided by Kayser et al. (2009) were two office environments, a courtyard, and a cafeteria, and were chosen to reflect situations often found in daily life. Realistic acoustic background noise was also available for each of the environments created by making binaural recordings in the environments as they were used by members of the public.

Each binaural impulse response could be convolved with a recording of a sound source to create a binaural stimulus that, when played back over headphones, would simulate the listening experience of listening to that sound source in one of the environments at one of the available listening positions within that environment. For the simulations built for the current study, the binaural impulse responses were used to place four different talkers around the listener within the cafeteria environment recorded by Kayser et al (2009). The talkers were positioned within the virtual environment as if they were sat at a table (positions A, B, and C in Figure 1). In the SiS condition, the talkers were present with no background noise. In the SiS+B condition, an excerpt from a continuous recording of natural café background noise recorded by Kayser et al (2009) in the same cafeteria environment was played continuously.

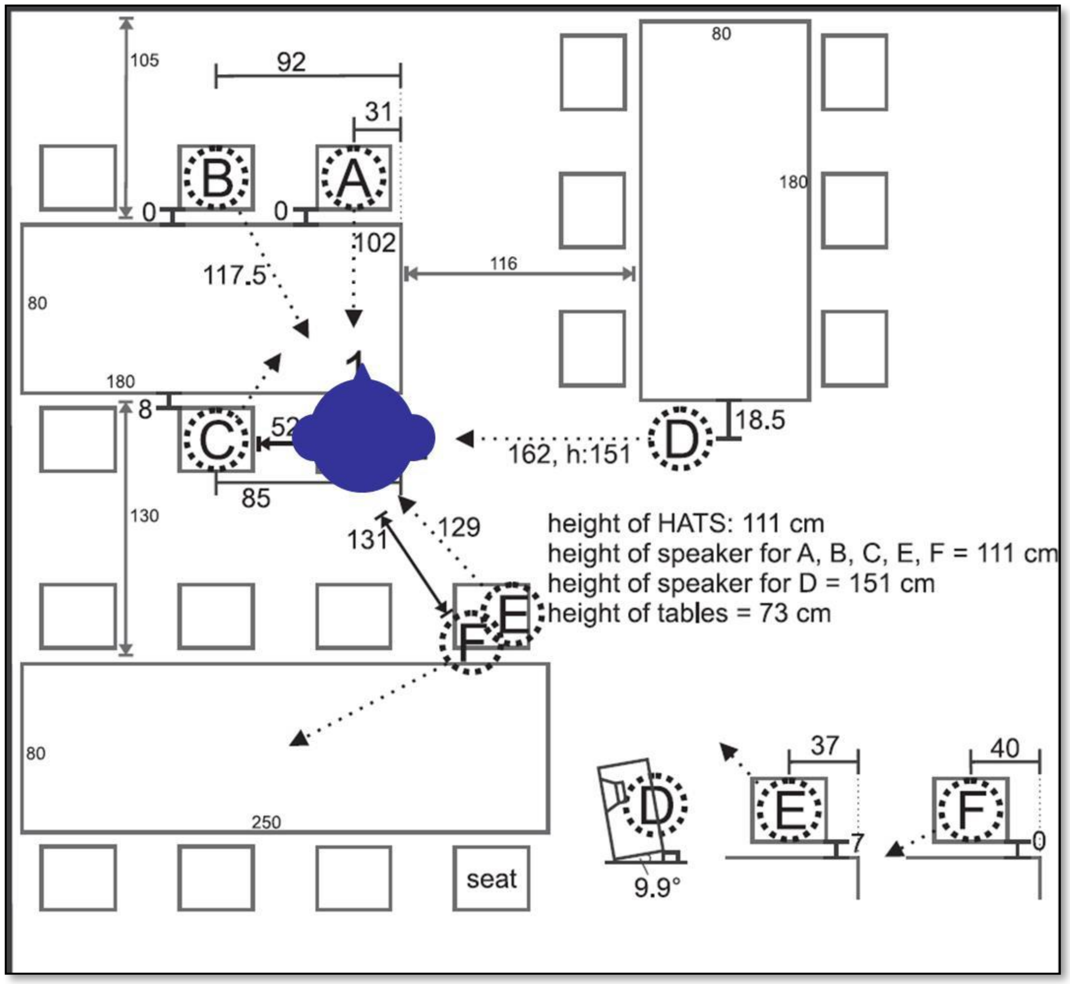


Figure 1: Graphical representation of the simulated listening environment (Kayser et al 2009). The letters represent simulated positions that sentences could be played from using the Oldenburg corpus of Head-Related Impulse Responses. Talkers were positioned at locations A, B, and C in both the SiS and SiS+B simulations. The blue head represents the location of the listener. The background noise used in SiS+B simulations had been recorded binaurally in this cafeteria environment when in everyday use and therefore captured realistic background noise originating from all directions.

The talkers spoke sentences from a version of the Co-ordinate Response Measure (CRM) corpus (Moore, 1981, Bolia et al., 2000) that had been recorded by male and female talkers with British and Irish accents (Kitterick et al., 2010). The CRM sentence corpus was identified as an acceptable stimuli of high quality, spoken by a range of different talkers in order to create realistic conversational interchange, and from a corpus that had a large number of unique sentence recorded so that it was not necessary to repeat sentences within or between simulations. Finally, as the CRM sentences followed a rigid format “Ready CALL-SIGN go to COLOUR NUMBER now”, for example “Ready Arrow go to blue one now”, this ensured that a listening task could be conducted using aspects of the sentences (such as identifying the colour said, number said etc.) meaning it was possible to quantitatively measure how well participants could understand speech in these situations with and without SSD.

Just as the spatial arrangement of the talkers were chosen to simulate a group conversation, the order in which the talkers spoke was also designed by the researcher to mimic a conversation (Figure 2). The talker seated in position A started to speak a new sentence after 1, 4, and 8 seconds, the target talker alternated between a male and female voices between simulations. As the average length of a CRM sentence was 2.5 seconds, this timing ensured that each new sentence that was spoken started after the previous one had ended. A separate back-and-forth exchange was simulated between the talkers seated in positions B and C by arranging the sentences such that each spoke after the other but the sentences that they spoke did not overlap entirely, again these talkers alternated between male and female voices but did not duplicate the target talker voice.

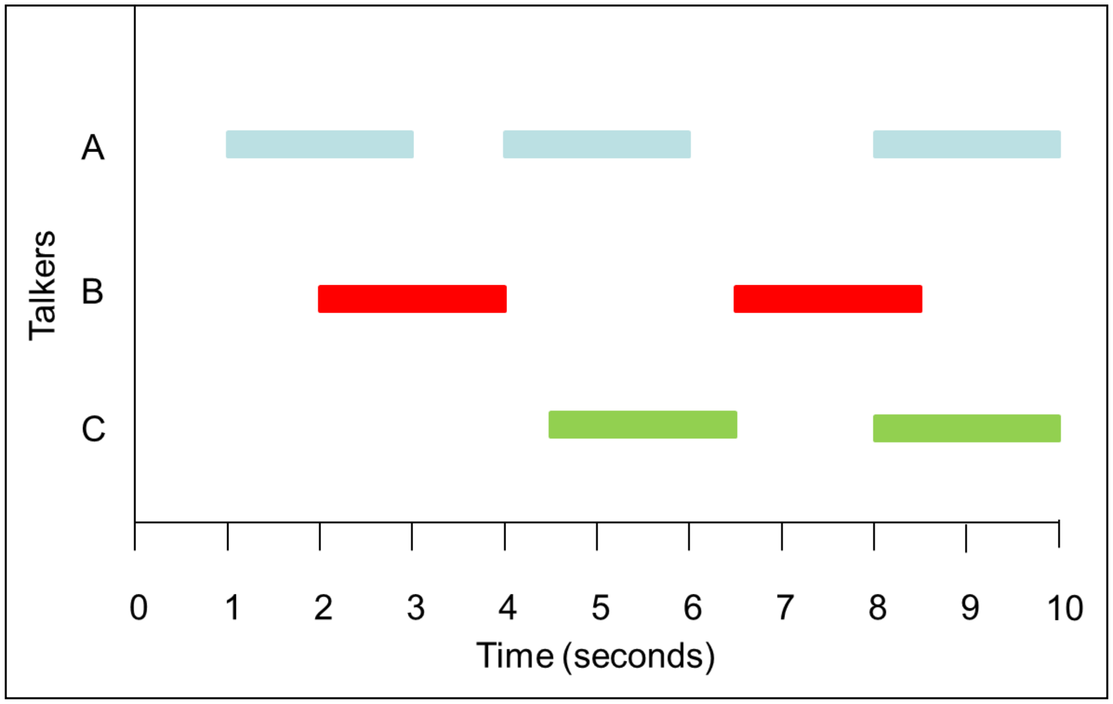


Figure 2: Graphical representation of the temporal configuration of talkers in the simulated conversation. Each colour bar represents the presentation of a sentence from the Co-ordinate Response Measure (CRM) corpus. The spatial positions of A, B, and C are shown in Figure 1.

3.3 Pilot 1

3.3.1 Introduction.

As using acoustic simulations of health states appears to be a novel but unproven method for describing a hearing-related health states, background work was required before proceeding to conducting the main experiments. Therefore, pilot testing was performed prior to the implementation of the main study in order to explore the feasibility of using ASH as a method of describing health and to provide data from which to base sample size calculations for the main experiments. In particular, the pilot experiment aimed to examine whether there was early evidence that ASH could communicate the fact that people with hearing loss are seen as having less than perfect health, which would be indicated if participants judged the hearing health states as having a utility value less than 1.0 (the utility value assumed to represent perfect health).

3.3.2 Methods

3.3.2.1 *Participants*

This experiment was conducted in accordance with the permissions granted by the NHS Grampian Research Ethics Committee and the Sponsor (University of Nottingham). Written informed consent was obtained for all participants by the Principal Investigator. Nine researchers based at Nottingham BRC participated in the pilot experiment. Demographic data were not collected. No financial incentives were given for participation in this pilot study. The sample size was chosen based on convenience to address initial feasibility questions and was not chosen to achieve sufficient statistical power for a specific statistical comparison. Participants were age 18 years or older, had self-reported normal hearing, and were willing and able to complete a short experiment involving the presentation of sounds, sounds and text or text only. Participants were asked to complete a TTO and VAS task (see section 3.3.2.2) based on how they thought the described hearing health state would affect their quality of life.

3.3.2.2 *Design: Valuation methods*

The time-trade off (TTO) technique and visual-analogue scales (VAS) were used as the health valuation methods (see Chapter 2 & 5 for a detailed description of both methodologies). The TTO approach presents a hypothetical situation to the individual considering the health state. They are informed that they will live in the health state described for a certain number of years and then they will die a painless immediate death after that time. The person is then given the choice of giving up years off the end of their life to return to a state of perfect health for the remaining years (Gudex, 1994). The trading tariff length chosen for this experiment, and all subsequent experiments, was ten years. Literature suggests that most studies use medium-length tariffs of ten or twenty years for TTO tasks (Attema et al., 2013), and the researchers thought that a ten-year tariff would be most relatable to all participants, regardless of their age and personal circumstance. Research has also suggested it is best to use a medium tariff length as TTO values tend to artificially decrease with an increased time frame (Arnesen and Trommald, 2005). Often differences in TTO values collected for different hearing-related health states are small (Summerfield et al. 2002) and therefore use of a larger tariff length (>10 years) could have meant that differences in TTO values between different presentation methods (e.g. text descriptions versus acoustic simulations) would be lost or over looked. However, the trading period was made to be adjustable in 6-month increments to acknowledge that differences in the TTO values of the different hearing-related health states were likely to be small.

The use of VAS to rate a health state is not a preference-based method of collecting health utility values. Instead, it provides a simple and straightforward way for participants to express an opinion about a health state. Comparison of the ranking of different health states using both VAS and TTO can be useful in determining whether participants had difficulty completing the TTO task (Gudex, 1994). The VAS used had endpoints represented by the labels ‘worst quality of life you can imagine', assigned the numerical value 0, and 'best quality of life you can imagine', assigned the numerical value 100.

3.3.2.3 *Design: Text based descriptions.*

Excerpts of vignettes developed by Lucas and Kitterick (2014) were used as the text-based description of living with profound SSD in the two situations of interest, SiS and SiS+B. The text descriptions used are detailed in Figure 3 and Figure 4. A limitation of using simulations to describe hearing-related health states is that the simulations can only recreate physical aspects of the hearing loss, such as a lack of access to binaural cues that create difficulties with localising sound, and do not recreate the emotional aspects of having a hearing loss, such as the frustration or exhaustion it can cause (Heffernan et al., 2016). Therefore, excerpts from the full text descriptions were used that pertained only to the physical limitations that an individual with SSD would experience and omitted any statements that related to the emotional impact of hearing loss.

- You lost all hearing in your left ear several years ago. You have no hearing loss in your right ear. You feel very upset if you think about losing the hearing in that ear.
- You cannot identify where sounds are coming from. All sounds appear to come from your right side.
- You can have a conversation in a small group as long as only one person talks at a time. You find it harder to follow the conversation if you cannot see everyone.
- You are often exhausted at the end of the day.

Figure 3: Text-based Descriptions of the SiS SSD-related health states.

- You lost all hearing in your left ear several years ago. You have no hearing loss in your right ear. You feel very upset if you think about losing the hearing in that ear.
- You cannot identify where sounds are coming from. All sounds appear to come from your right side.
- You try to avoid places with lots of background noise such as pubs or cafés. Having a conversation in these places can be difficult, frustrating, and isolating.
- You are often exhausted at the end of the day, especially after listening in background noise

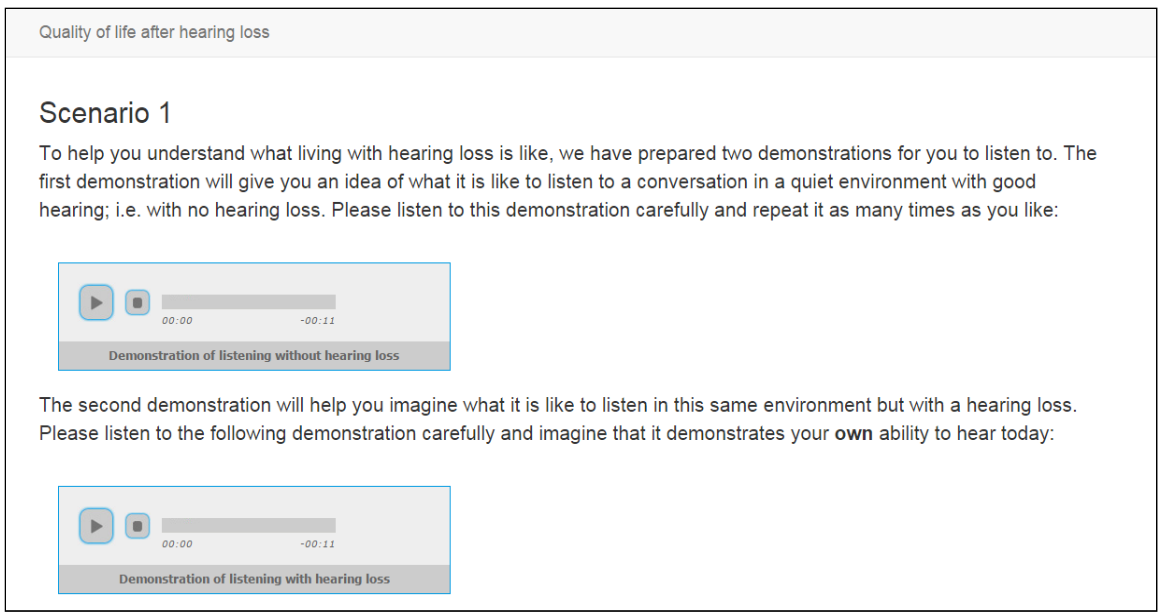
Figure 4: Text-based Descriptions of the SiS+B SSD-related health states.

3.3.2.1 *Design: Web-based interface*

A purpose-built web-based interface was used to present all study activities to participants via a web-page (accessed via google chrome), excluding written informed consent which was obtained using paper forms. The webpage was programmed to present simulations one at time with a TTO and VAS task appearing after the participant had listened to the simulation (see figures 5, 6, 8, and 9). Specific details of the procedure administered using the web-based interface are given in Section 3.3.2.3. The participant's answers were collected and stored by the web interface locally on the study laptop, not via the internet, to ensure the data were stored safely. However, the application was designed to have the capability to collect valuations remotely for future experiments. The web-based interface was used to present all conditions.

	Block 1	**Block 2**
Modality:	Audio only (A alone)	Audio and Text (AT) vs Text only (T)
Listening Environment:	SiS SiS+B	

Figure 5: Experiment outline of the pilot conditions.



Fi

Figure 6: An example of an audio only presentation of the hearing-related health state.

3.3.2.2 *Design: General*

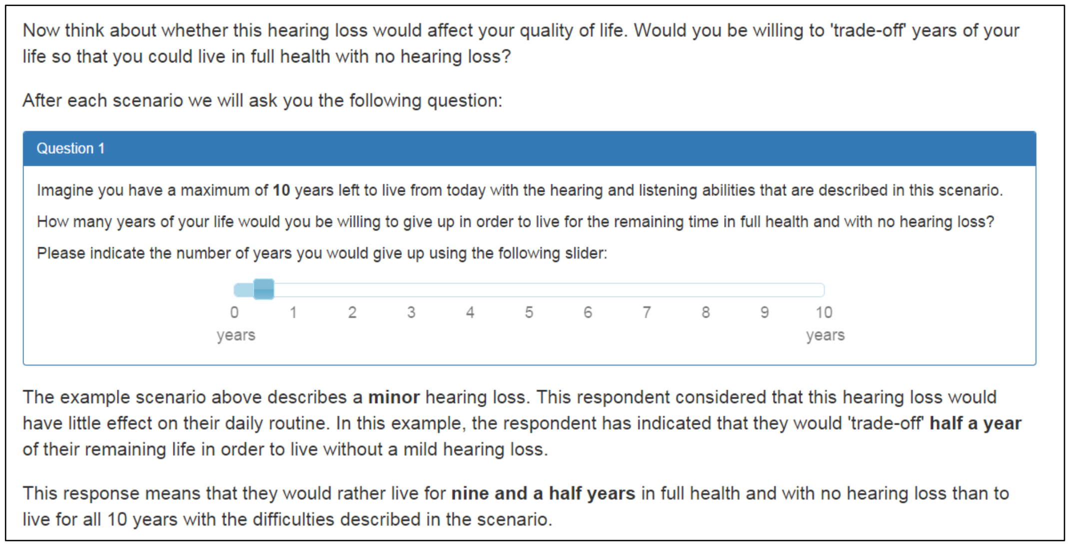
Six conditions were defined from the factorial combination of the three possible presentation modalities (Audio only, Audio and Text, and Text only; A, AT, and T, respectively) and the two listening environments (SiS, SiS+B). Each participant took part in a single session in which the conditions were presented within separate blocks, completed in a fixed order (Figure 5). The ‘Audio alone’ (A alone) conditions were presented in the first block to ensure that the participants had not already read the text descriptions and therefore that their responses were based on the information conveyed by the acoustic simulations alone (see Figure 6 for presentation of the ASH). The conditions that included the text descriptions (AT and T) for both the Sis and SiS+B listening environment were presented in a random order in the second block. All data were collected electronically and stored in line with General Data Protection Regulations and analysed using SPSS, this procedure was followed for all subsequent experiments within this body of research.

3.3.2.3 *Procedure*

Simulations were calibrated using a Bruel and Kjær Type 2250 sound pressure level meter which was attached to an artificial ear in which a pressure field ½“ type 4192 microphone was used. The sound level meter was calibrated against a 94dB sound pressure level 1 kHz tone reference, type 4231. The Beyerdynamic DT 770 Pro Studio headphones were placed over the artificial ear and ten simulations were played (see figure 2), the average sound level was taken across all simulations. Simulations were presented at 65.1 dB in the left ear, normal hearing ear, and 0 dB in the ear with simulated profound deafness. Simulations were presented at 64.7 dB, within the 1 dB tolerance level set by the researcher. A level of 65 dB was chosen as this is the sound level for normal speech (Elberling & Worsoe, 2006).

Figure 7: Set up of artificial ear for calibration of sound level of simulations. Top photograph demonstrates the artificial ear and microphone used. Bottom photograph shows the headphones placed over the artificial ear during sound calibration.

Participants were seated at a computer, which was situated in a double-walled sound-attenuated room and asked to put on a pair of closed back headphones (Beyerdynamic DT 770 Pro Studio) so that they comfortably enclosed their outer ears. Participants were first presented with examples of the TTO and VAS exercises for health states not related to SSD. The examples described someone who had a stroke and someone who had mild hearing loss. They were then presented with A alone SiS simulation and a version of the same simulation but with normal hearing, and A alone SiS+B and a version of the same simulation but with normal hearing (Block 1) and asked to value each state using TTO (Fig. 7) and VAS (Fig. 8) tasks. Participants were then presented with in AT and T simulations of normal hearing, SiS, and SiS+B, and asked to value each state using a VAS (figure 7) and TTO task (figure 8). Participants were asked to complete a health valuation exercise after each description was presented using VAS and TTO tasks. The experiment session lasted approximately 20 minutes.


Now think about whether this hearing loss would affect your quality of life. Would you be willing to 'trade-off' years of your life so that you could live in full health with no hearing loss?

After each scenario we will ask you the following question:

Question 1

Imagine you have a maximum of **10** years left to live from today with the hearing and listening abilities that are described in this scenario.

How many years of your life would you be willing to give up in order to live for the remaining time in full health and with no hearing loss?

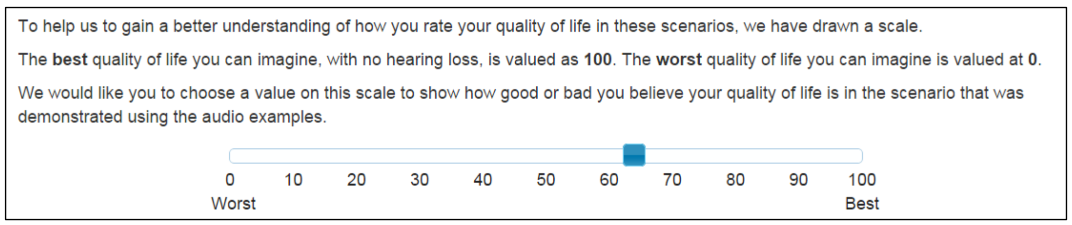
Please indicate the number of years you would give up using the following slider:

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

The example scenario above describes a **minor** hearing loss. This respondent considered that this hearing loss would have little effect on their daily routine. In this example, the respondent has indicated that they would 'trade-off' **half a year** of their remaining life in order to live without a mild hearing loss.

This response means that they would rather live for **nine and a half years** in full health and with no hearing loss than to live for all 10 years with the difficulties described in the scenario.

Figure 8: Pilot study example TTO slider


To help us to gain a better understanding of how you rate your quality of life in these scenarios, we have drawn a scale.

The **best** quality of life you can imagine, with no hearing loss, is valued as **100**. The **worst** quality of life you can imagine is valued at **0**.

We would like you to choose a value on this scale to show how good or bad you believe your quality of life is in the scenario that was demonstrated using the audio examples.

0 Worst 10 20 30 40 50 60 70 80 90 100 Best

Figure 9: Pilot study example VAS slider

3.3.2.1 *Analysis*

VAS data were for Pilot 1 was transformed into utility values (U VAS) using the formula previously detailed in Chapter 2 (see section 2.5.3). The TTO data set was automatically transformed into a utility value by the web interface, using the formula also detailed in Chapter 2 (see section 2.5.2). As the pilot study was not statistically powered only descriptive statistics were performed on the TTO and U VAS values.

3.3.3 Results

The average transformed VAS ratings for each condition are presented in Table 2. On average, the SiS was rated higher on the VAS scale than the SiS+B regardless of the modality of the information that was provided to describe both listening situations. Informed by just the acoustic simulations, six out of the nine participants (66%) rated the SiS health state simulation as better health than the SiS+B health state simulation and three gave identical ratings. None of the participants rated the SiS+B health state simulation as healthier than SiS. A similar number of participants rated SiS health state higher than SiS+B health state on the VAS scale when presented with both text only and the combination of audio and text.

For the TTO task no participant was willing to trade-off more years of life in the SiS situation compared to the SiS+B situation. The average valuation of the SiS health state simulation was also higher than that of the SiS+B health state simulation, regardless of the modality of the information that was provided; i.e. audio only, text only, or the combination of both. Non-trading rates appeared to be slightly higher when only audio information was provided compared to both text only and both modalities together. However, there was evidence that participants were still able and willing to trade off years of life based on just the acoustic information alone.

Participant	Audio alone (A)		Text only (T)		Audio and Text (AT)	
	SiS	**SiS+B**	**SiS**	**SiS+B**	**SiS**	**SiS+B**
1	0.99	0.97	0.99	1.00	1.00	0.97
2	0.92	0.85	0.96	0.92	0.89	0.92
3	0.98	0.97	0.95	0.96	0.95	0.94
4	0.97	0.97	0.96	0.67	0.92	0.95
5	0.97	0.97	0.92	0.88	0.89	0.88
6	0.99	0.99	0.99	0.98	0.99	0.98
7	1.00	1.00	0.85	0.85	0.85	0.85
8	0.97	0.92	0.92	0.92	0.97	0.92
9	0.99	0.98	0.97	0.92	0.97	0.97
Average	**0.98**	**0.97**	**0.94**	**0.92**	**0.95**	**0.94**

Table 2: Transformed U VAS data for Pilot 1 for listening conditions SiS (speech in speech), SiS+B (speech in speech with background noise) and presentation methods

Participant	**Audio only (A)**		**Text only (T)**		**Audio and Text (AT)**	
	SiS	**SiS+B**	**SiS**	**SiS+B**	**SiS**	**SiS+B**
1	0.95	0.95	0.95	0.95	0.95	0.90
2	0.85	0.75	0.90	0.85	0.85	0.80
3	1.00	1.00	0.95	0.95	0.95	0.95
4	1.00	1.00	1.00	1.00	1.00	1.00
5	1.00	1.00	1.00	1.00	0.95	1.00
6	0.95	0.95	0.95	0.95	0.95	0.95
7	1.00	1.00	0.80	0.80	0.80	0.80
8	0.95	0.95	0.95	0.95	0.95	0.95
9	0.95	0.90	0.90	0.75	0.90	0.85
Average	**0.96**	**0.94**	**0.93**	**0.91**	**0.92**	**0.91**

Table 3:TTO data for Pilot 1 listening conditions SiS (speech in speech), SiS+B (speech in speech with background noise) and presentation methods

3.3.4 Discussion

The results of the pilot testing appear to indicate that using ASH to communicate the fact that people with hearing loss are seen as having less than perfect health could be feasible. The utility values obtained using ASH to simulate a hearing-related health state were less than one, and listening environments had an effect, with average SiS+B health utility values lower than those for SiS as across modalities; audio, audio and text and text only. The effect of presentation modality was not clear, so the second hypothesis remains unanswered.

The pilot study had some shortcomings. Firstly, the pilot participants were mainly hearing researchers, and this could have influenced the utility values obtained as the hearing researchers could have had a different view of the health state than, for example, members of the general public. As discussed by Ubel et al. (2003), there is often a difference between the valuations obtained from patients and the general public, and this difference is likely due to the different life views of these groups. The public are likely viewing any given health state in terms of the loss of good health, whereas patients or those with topic expertise about the state are possibly looking at the potential to regain health in the future. To avoid this potential confound in future studies, members of the general public were asked to give valuations of the health states and also asked about their knowledge of hearing loss to help to identify the reasons for any potential outliers within future experiments.

Secondly the pilot testing involved passively listening to audio clips. It is possible that as participants did not have to actively engage with the simulations they did not perceive the difficulty with sound localisation that was being conveyed as they did not need to determine where a talker was speaking from and therefore did not receive feedback on their (in)ability to perceive their location. If so, they may has assumed that as they could still hear sounds then their quality of life would not be affected significantly by the hearing loss being simulated.

Furthermore, verbal feedback collected from the pilot participants regarding the simulations and the procedure suggested that modifications to the simulations were required. Overall, participants felt that the simulations sounded very unnatural and that the CRM sentences stood out from the cafeteria background noise. An alternative sentence corpus was therefore used in further versions of simulations, and these changes to the simulations are detailed in section 3.3.5 below. Verbal feedback also suggested that changes to the web-based interface used to obtain the valuations would be helpful to capture participants' preferences accurately. Specifically, participants indicated that they would have preferred being able to adjust the amount of time traded-off in smaller increments. The time period to be traded could only be adjusted six-month intervals. It may also have been preferable for the VAS question to be hidden from view until the TTO question is completed, as being able to view the VAS response scale may have anchored subsequent responses.

3.3.5 Alterations made to version 1 simulations

Changes were made to the set-up of the simulations with the talkers now speaking more naturalistic sentences in a less rigid format. No changes were made to the simulation environment or simulation set up; all simulations were still situated in the cafeteria environment. As with version 1 of the simulations the sentences presented from speaker A position were sentences from the IEEE sentence corpus (1969) at 2, and 10 seconds approximately. During this time there was a back and forth conversation between speakers B and speakers C, the sentences presented from speaker B, and C position were sentences from the Bamford-Kowal-Bench (BKB) sentence corpus (Bench et al., 1979). Talker D was removed from the version 2 simulations after feedback from participants that the interjection felt unnatural. The BKB sentence-corpus consists of 2 distinct female speakers and 2 distinct male speakers with English accents. The sentence corpus uses more naturalistic sentences than the CRM sentences, for example, “the clown had a funny face” or “the park’s near the road”. Figure 10 is a graphical representation of the temporal arrangement of the talkers. To alleviate any possible systematicity in *when* a talker started to speak, and to make the simulations more realistic, a random offset was applied to the onset time of each sentence so that each talker did not start to say a sentence at the exact same time in each simulation, making the simulations less predictable.

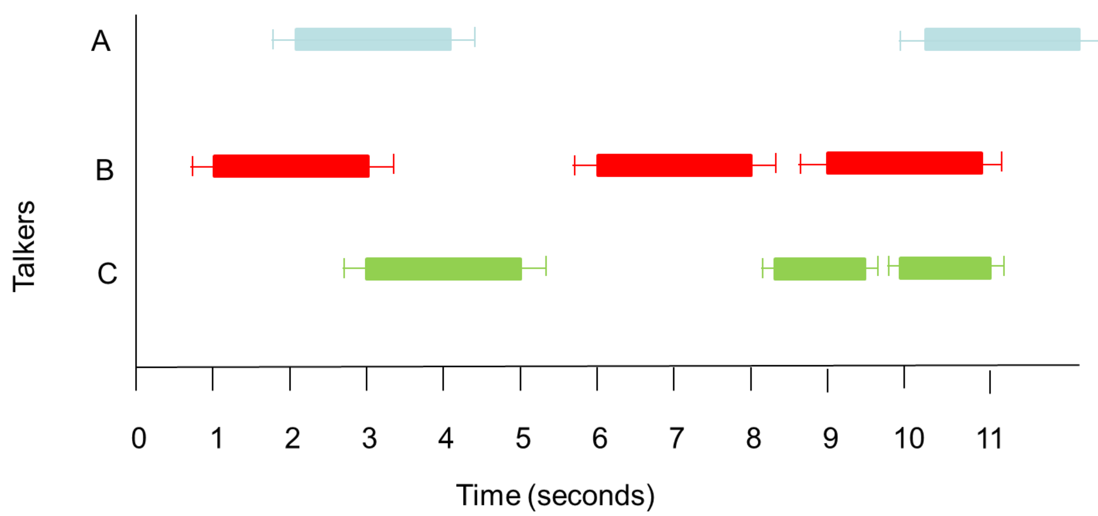


Figure 10: Graphical representation of the temporal configuration of talkers in the simulated conversation conveyed in the revised simulations. Each colour bar represents the presentation of a new sentence. The colour blocks represent sentences and tails on either side of each coloured block represent the range over which the onset time was varied in order to reduce the predictability of the simulated conversation.

3.4 Experiment 1

3.4.1 Introduction

Pilot testing indicated that the use of ASH when describing hearing-related health appears feasible. However, several questions would need to be addressed before ASH could confidently be used as a method of describing hearing-related health states within health state valuation exercises. Firstly, it is necessary to determine whether ASH are capable of conveying different levels of listening difficulty, in the same way that a text-based description is capable of communicating different degrees of severity. One possible approach would be to include an ASH of ‘healthy hearing’ (i.e. no hearing impairment) alongside the SiS and SiS+B conditions piloted previously that convey the difficulties associated with single-sided deafness. The hypothesis being that the average utility values for simulations of SSD would be expected to be lower than the average utility scores obtained from a ‘healthy hearing’ ASH, just as utility values of hearing-impaired health states described through traditional methods (i.e. text-based descriptions) have been found to be significantly lower than perfect health (Summerfield et al., 2002, Summerfield et al., 2010). If the ASH is successful in describing a hearing-related health state, then utility values obtained using ASH should be valued as less than perfect health.

Secondly, as simulations have not been used for the purpose of describing health states related to hearing loss before, the validity and robustness of the impact of ASH require investigation. For example, are valuations obtained using ASH easily influenced by information from other sources, such as previous exposure to a related text-based description? To assess this question, it would be necessary to elicit preferences for a health state described using ASH alone, and then to expose participants to a related text-based description and then re-assess their preference for the sample health state. A further consideration is whether providing a combination of ASH and text simultaneously influences utility values over providing either alone. It was hypothesised that using both mediums to describe health states related to SSD would result in lower (poorer) mean utility values than when text alone is presented as the participants will have an applied example of what it’s like to hear with a hearing loss in addition to an in-depth written description of the effects of living with a hearing loss. Finally, it is necessary to assess whether valuations obtained using ASH alone or a combination of ASH and text are as stable as valuations collected using the traditional gold standard text-based descriptions. ASH cannot be a viable method if it is very unstable and requires multiple presentations in order to gain stable valuations.

3.4.1.1 *Summary of research questions*

Experiment one aims to address the following research questions:

1. Can Audio alone (A) descriptions of hearing-related health states communicate the fact that people with hearing loss are seen as having less than perfect health?
2. Can Audio alone (A) descriptions of hearing-related health states communicate different levels of listening difficulty?
3. What effect is there on health state valuations when the health state is described using Audio and Text (AT) vs Text alone (T)? The hypothesis being that using A+T to describe health states related to SSD will result in lower mean health valuations than using T alone.
4. Are valuations obtained using A alone susceptible to influence from information gained from other sources; i.e. from text-based descriptions provided during the experiment?
5. Are valuations collected using A alone or AT stable compared to valuations collected using T alone?

6. Are valuations collected using A alone related to valuations collected using T and are valuations collected using AT related to valuations collected using T?

3.4.2 Methods

3.4.2.1 *Participants*

This experiment was conducted in accordance with the permissions granted by the NHS Grampian Research Ethics Committee and the Sponsor (University of Nottingham). Written informed consent was obtained for all participants by the Principal Investigator. Using data collected from pilot 1 an *a priori* power analysis was computed to detect within-subject differences in the mean health state valuation at alpha=0.05 and a power of 0.80. The required sample size varied between 8 (based on the effect size of 1.04 SD for A alone SiS group and text alone SiS group) and 16 (based on the effect size of 0.62 SD for A alone SiS+B group and text alone SiS+B group). Therefore, sixteen normal hearing participants (7 males, 9 females) who were naïve listeners and not hearing specialists took part in Experiment 1. Participants had an age range of 18 to 76 years old (mean age 33.4 years) and were recruited from the National Institute for Health Research Nottingham Biomedical Research Centre (NIHR Nottingham BRC) participant database, and from posters placed in student areas on the University of Nottingham University park campus. Participants were naïve listeners deliberately chosen so as not to include hearing researchers or individuals with specific expertise about hearing. All participants had pure-tone average thresholds between 0.5 and 4 kHz ≤ 30 dB HL [Hearing Level] (i.e. passed a screen for normal hearing) in both ears and did not have a current ear infection or any other condition that would prohibit a hearing assessment taking place prior to the first testing session Participants were financially compensated for their time at a rate of £5 for up to one hour of participation and reimbursed for any incurred travel expenses. The experiment lasted approximately 45 minutes..

3.4.2.2 *Design: valuation methods*

Pilot participants indicated that they would have preferred the ability to adjust the amount of time traded-off in smaller increments. For Pilot 1, the period of time that could be traded was set to six-month intervals. For Experiment 1, the TTO trade off sensitivity was set at monthly increments and the increments were clearly displayed on screen (Figure 9). The TTO response was removed from view when the VAS appears to avoid the TTO response anchoring respondents' subsequent decisions or vice versa.

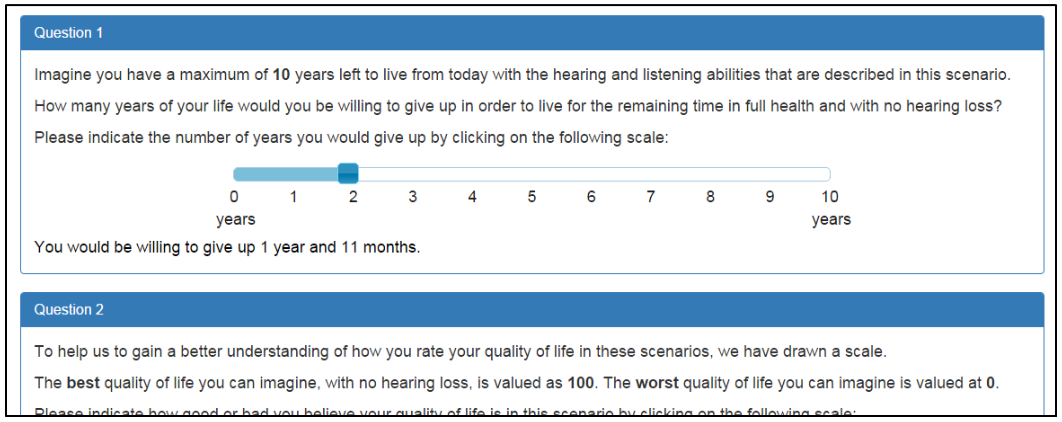


Figure 11: An example of how the TTO trade off sensitivity was increased by permitting the period traded off to be defined in 1-month steps, with visual and written feedback on the period that was selected by the participant.

3.4.2.3 *Design: text-based descriptions*

The text-based descriptions of the listening environments remained unchanged from Pilot 1. However, three reference health states were added to the experiment. These reference health states were taken from the Health Utilities index Mark III (HUI 3) (Feeny et al., 2002) (for an examples, see Figure 11). The reference states were included to confirm that participants did indeed view perfect health as having a health utility value of 1 or close to one, and that participants viewed the health utility of an individual with deafness in line with previously collected utility values of approximately 0.5 (Cheng and Niparko, 1999). A comparison between the utility values assigned to these reference health states and those obtained using the acoustic simulations would also confirm whether the simulations communicated that people with hearing loss are seen as having less than perfect health.

Scenario 1

Please read the following text imagining that it describes your **own** health today.

Able to see well enough to read ordinary newsprint and recognize a friend on the other side of the street, without glasses or contact lenses.

Able to hear what is said in a group conversation with at least three other people, without a hearing aid.

Able to be understood completely when speaking with strangers or friends.

Able to walk around the neighbourhood without difficulty, and without walking equipment.

Full use of two hands and ten fingers.

Happy and interested in life.

Able to remember most things, think clearly and solve day to day problems.

Free of pain and discomfort.

Perfect health (healthy) HUI3 description

Scenario 2

Please read the following text imagining that it describes your **own** health today.

Able to see well enough to read ordinary newsprint and recognize a friend on the other side of the street, without glasses or contact lenses.

Unable to hear at all.

Able to be understood completely when speaking with strangers or friends.

Able to walk around the neighbourhood without difficulty, and without walking equipment.

Full use of two hands and ten fingers.

Happy and interested in life.

Able to remember most things, think clearly and solve day to day problems.

Free of pain and discomfort.

Perfect health with deafness (deaf) HUI3 description

Scenario 3

Please read the following text imagining that it describes your **own** health today.

Unable to see at all.

Unable to hear at all.

Unable to be understood when speaking to other people (or unable to speak at all).

Cannot walk at all.

Limitations in use of hands or fingers, requires the help of another person for all tasks (not independent even with use of special tools).

So unhappy that life is not worthwhile.

Unable to remember anything at all, and unable to think or solve day to day problems.

Severe pain that prevents most activities.

Worst health imaginable (worst) HUI3 description

Figure 12: Reference health state descriptions defined using the descriptive system of the HUI3 I

3.4.2.4 *Design: general*

The experiment included six unique hearing-related descriptions resulting from a 2x3 factorial design with three levels for the modality of health state presentation (A, A+T and T) and two levels of listening environment (SiS, SiS+B). The main valuation task comprised 3 testing blocks (Figure 13). In Block 1, health states were described in the SiS and SiS+B listening environments using acoustic simulations alone (A); i.e. with no text description. As in Pilot 1, this design ensured that these initial valuations were not influenced by knowledge of the associated text descriptions of the same situations. Block 2 contained all conditions that included the presentation of the text description (Audio and Text (AT) and Text (T) alone) presented in a counterbalanced order. Each of the four conditions in Block 2 were repeated twice. Finally, in Block 3, the health states described by presenting A alone were valued for a second time. The reference health states from the HUI3 were valued at the start and end of the experiment; i.e. before Block 1 and after Block 3.

By presenting audio alone in Block 1, it was possible to address the question of whether acoustic simulations alone can communicate the fact that people with hearing loss are seen as having less than perfect health. If using of A alone results in utility valuations that are less than perfect health (as quantified by the HUI3 reference health states), then the ASH may be feasible method of describing hearing-related health states. Presenting AT and T in Block 2 addressed the second and third research questions; i.e. what effect is there on health state valuations when the health state is described using A+T vs T. Having AT and T valuation data for the SSD-related health states makes it possible to compare the different modalities and the different listening environments. The valuation data collected using A alone in block 1 and AT and T in Block 2 can also be compared and the effect, if any, of modality and listening environment on valuation data can be assessed.

Presenting ASH alone in block 1 and then again in block 3 allows for the initial utility values described using ASH alone (collected prior to any familiarity with the text-based description) to be directly compared to utility values described using ASH alone *after* the participant has gained supplementary knowledge from the text descriptions presented using AT and T in block 2. If the valuations in block 1 and 3 differ, it is plausible to hypothesise that this could be caused by the prior knowledge that participants have gained from text descriptions previously presented in block 2; i.e. that the valuations obtained using A alone are susceptible to influence from information gained from other sources.

	Block 1	Block 2	Block 3
Modality:	Audio only (A alone)	Audio and Text (AT) vs Text only (T)	Audio only (A alone)
Listening Environment:	SiS vs SiS+B (Counterbalanced)		

Figure 13: Experiment outline of the three main blocks of the valuation task. The conditions in which the health states were described using acoustic simulations alone were presented in Blocks 1 and 3. Block 2 contained health states described either using text or the combination of text and an acoustic simulation.

3.4.2.1 *Counterbalancing procedure.*

At the start of the experiment, the reference health states were presented in a fixed order: perfect health, perfect health with complete deafness, and worst health imaginable. They were then presented in the reverse order at the end of the experiment. Counterbalancing was carried out within each experimental block. In Block 1, SiS was presented first for half of the participants and SiS+B was presented first for the other half. They were then presented in the opposing order when repeated in Block 3. In Block 2, the AT and T presentations of SiS and SiS+B were counterbalanced using a Latin square design across participants to minimise order effects.

3.4.2.1 *Calibration of experiment using artificial ear*

Simulations were calibrated using a Bruel and Kjær Type 2250 sound pressure level meter which was attached to an artificial ear in which a pressure field ½-inch type 4192 microphone was used. The sound level meter was calibrated against a 94-dB Sound Pressure Level (SPL) 1 kHz tone reference (Bruel and Kjær Type 4231). Beyerdynamic DT 770 Pro Studio 80 ohm closed-back headphones were placed over the artificial ear and ten simulations were played, and the average sound level was taken across all simulations. Simulations were presented at 65 dB SPL in the left (normal hearing) ear alone. A level of 65 dB SPL was chosen as this approximately the sound pressure level for normal conversational speech. Presentation levels were calibrated to be within 1 dB of the target level.

3.4.2.1 *Procedure*

Participants gave informed consent to participate in the study and to consent to a hearing screening test. Once the consent was obtained, a screening audiogram was carried out by the researcher to ensure participants had a pure-tone average (PTA) threshold of ≤ 30 dB HL [Hearing Level] between 0.5 and 4 kHz (Audiology, 2011).

Participants were seated at a computer which was situated in a double-walled sound-attenuated room and asked to put on a pair of Beyerdynamic DT 770 Pro Studio closed-back headphones so that they comfortably enclosed their pinnae. Participants were then presented with examples of the TTO and VAS exercises for health states not related to hearing loss to ensure they understood the principles of the TTO and VAS tasks. Participants were asked to complete a health valuation exercise after each description was presented using both VAS and TTO tasks. Once the initial testing session was completed, participants returned for a second testing session to repeat the entire experimental procedure again to assess the stability of scores across test sessions.

3.4.2.2 *Analysis*

The data from Experiment 1 relating to aims 1-4 were analysed using paired t-tests and aims 4-5 relating to the reliability of the utility values were analysed using intra-class coefficients (ICC). ICC analysis is the recommended method for test-retest reliability (Altman and Bland, 1983, Bland, 2015). Reliability is expressed as the proportion of variance that can be attributed to the true score:

$$ICC_{agreement} = \frac{\sigma_S^2}{\sigma_S^2 + \sigma_{time}^2 + \sigma_{err}^2}$$

This formula is the basis for all intra-class correlations: the magnitude of the reliability coefficient is related to the variability of scores between individuals, the true score, ($\sigma 2$ s) and total variance, that is the variability between individuals, and measurement error ($\sigma 2$ err).

ICC values are reported as a ratio of all variances, ranging from 0 (unreliable) to 1 (excellent reliability). They are sensitive to random error and systematic bias and are sample dependent, so a homogenous sample (one with closely grouped scores) will have less variability and the magnitude of the ICC will be smaller, indicating less reliability to differentiate (Altman and Bland, 1983, Bland, 2015, de Vet et al., 2011, Griffiths and Murrells, 2010). Many researchers have recommended that confidence intervals for the ICCs are reported alongside an ICC value to help determine the accuracy of the ICC estimates (Shrout and Fleiss, 1979, Atkinson and Nevill, 1998, DeVon et al., 2007, Bartlett and Frost, 2008, Streiner and Norman, 2008). As ICCs estimate the proportion of variance in observed scores that can be attributed to the true score, generally ranging from 0 to 1, an ICC estimate of 0.77 for example would indicate that approximately 77% of the variability in the data can be attributed to the true score (Weir, 2005, Streiner and Norman, 2008, Bland, 2015). There are many suggestions in the literature regarding the interpretation of the ICC values, and more specifically what level constitutes a high level of reliability. The minimum standards for test-retest estimates range anywhere from 0.60 to 0.95, depending on the purpose of the tool (Terwee et al., 2007, Streiner and Norman, 2008, de Vet et al., 2011, Nunnally, 1978, Shoukri et al., 2004). This research follows guidance given by Koo and Li (2016) which states that “values less than 0.5, between 0.5 and 0.75, between 0.75 and 0.9, and greater than 0.90 are indicative of; poor, moderate, good, and excellent reliability, respectively”.

Aim 6 looked at the correlations between the data sets using Pearson’s correlation coefficient, also known as Pearson’s r. Pearson’s r is a measure of the linear association; i.e. the strength of association between two variables. Pearson’s r values range from -1 (strongly negatively associated) to 1 (strongly positively associated), with 0 denoting no association between variables. A positive Pearson’s r value implies that an increase in the first variable would relate to an increase in the second variable and a negative Pearson’s r value implies that a decrease in the first variable would relate to an increase in the second variable (Howitt and Cramer, 2007, Mukaka, 2012, Taylor, 1990). Pearson’s r values have previously been categorised by researchers to assist the interpretation of correlations. In a review of correlation use in medical literature Taylor (1990) stated that a correlation of: ≤ ± 0.35 denote weak correlation, ±0.36 - ±0.67 denotes moderate correlation and ±0.68 – ±0.9 denotes high correlation and ≥ ± 0.9 denotes a very high correlation. More recently Mukaka (2012) categorised Pearson’s r values further stating that: ±0.0 - ±0.3 denotes weak correlation, ±0.3 - ±0.5 denotes low correlation, ±0.5 - ±0.7 denotes moderate correlation, ±0.7 – ±0.9 denotes high correlation and ≥ ± 0.9 denotes a very high correlation.

3.4.3 Results

The health utility results collected using TTO and averaged across the repeated presentations within each session are shown in Table 2 separately for session one and session two. Averages of these data across both sessions are shown in Table 3. The reference health states collected indicate that participants judgements were in line with expectations and that the TTO task appeared to be generally understood. The average utility value for perfect health was close to 1 (healthy = 0.95), the average score for perfect health with deafness was close to figures previously found in the literature (deaf = 0.65) and average score for worst health imaginable was close to 0 (worst = 0.06).

The first aim of this experiment was to determine if A alone descriptions of hearing-related health states can communicate the fact that people with hearing loss are seen as having less than perfect health. One way paired-samples t-tests were conducted to compare utility values for the healthy reference state and the A alone valuations of the SiS and SiS+B listening situations. There was a significant difference in average utility values collected for the healthy reference condition (M= 0.95, SD = 0.08) and the A alone SiS+B condition (M = 0.64, SD = 0.18) ($t(15)=6.71$, $p<0.01$).

The second aim of the experiment was to determine if Audio alone (A) descriptions of hearing-related health states communicate different levels of listening difficulty. A one way paired-samples t-test was conducted to compare utility values for A alone SiS condition and A alone SiS+B condition. There was a significant difference in average utility values collected for the A alone SiS (M= 0.73, SD= 0.15) A alone SiS+B (M= 0.65, SD= 0.19), $t(15)=4.26$, $p<0.01$.

The third aim was to assess the effect of describing the health state using Audio and Text (A+T) vs Text alone (T) on health state valuations. For SiS, there was no significant difference between when the health state is described using Audio and Text (A+T) (M=0.65, SD=0.17) vs Text alone (T) (M=0.65, SD=0.17) $t(15)=0.49$, $p=0.63$. For SiS+B, there was no significant difference between when the health state is described using Audio and Text (A+T) (M=0.62, SD=0.19) vs Text alone (T) (M=0.61, SD=0.19) $t(15)= 1.11$, $p=0.28$.

	Session 1 average utility values									Session 2 average utility values								
Participant	Healthy	Deaf	Worst	A SiS	A SiS+B	AT SiS	AT SiS+B	T SiS	T SiS+B	Healthy	Deaf	Worst	A SiS	A SiS+B	AT SiS	AT SiS+B	T SiS	T SiS+B
1	1.000	0.250	0.000	0.550	0.350	0.400	0.300	0.400	0.300	1.00	0.20	0.00	0.28	0.20	0.28	0.20	0.20	0.20
2	1.000	0.755	0.005	0.740	0.740	0.715	0.725	0.735	0.720	1.00	0.79	0.02	0.80	0.73	0.69	0.64	0.69	0.67
3	0.945	0.800	0.015	0.745	0.640	0.595	0.560	0.645	0.585	1.00	0.62	0.01	0.74	0.62	0.72	0.63	0.63	0.64
4	0.845	0.600	0.205	0.765	0.745	0.730	0.700	0.705	0.715	0.99	0.69	0.20	0.71	0.65	0.70	0.67	0.70	0.69
5	0.985	0.510	0.010	0.950	0.710	0.680	0.570	0.730	0.540	1.00	0.79	0.01	0.97	0.90	0.91	0.86	0.96	0.83
6	0.825	0.620	0.015	0.325	0.250	0.250	0.200	0.400	0.200	1.00	0.60	0.01	0.73	0.18	0.35	0.15	0.30	0.21
7	0.955	0.615	0.050	0.800	0.685	0.715	0.715	0.745	0.730	0.97	0.69	0.14	0.76	0.78	0.75	0.74	0.76	0.76
8	1.000	0.800	0.025	0.545	0.555	0.525	0.485	0.550	0.485	1.00	0.63	0.19	0.73	0.67	0.72	0.68	0.68	0.67
9	1.000	0.745	0.000	0.695	0.595	0.620	0.650	0.575	0.540	1.00	0.75	0.00	0.76	0.70	0.73	0.69	0.73	0.70
10	0.965	0.700	0.005	0.850	0.800	0.865	0.870	0.865	0.865	1.00	0.80	0.00	0.94	0.94	0.92	0.90	0.92	0.90
11	0.970	0.495	0.010	0.795	0.695	0.780	0.650	0.600	0.500	0.99	0.61	0.01	0.94	0.92	0.93	0.88	0.87	0.85
12	0.895	0.740	0.105	0.395	0.355	0.290	0.300	0.295	0.355	1.00	0.80	0.00	0.94	0.94	0.92	0.90	0.92	0.90
13	0.990	0.625	0.110	0.900	0.880	0.885	0.900	0.895	0.805	0.93	0.77	0.11	0.91	0.82	0.80	0.80	0.77	0.80
14	0.645	0.165	0.170	0.495	0.495	0.295	0.460	0.520	0.395	0.66	0.75	0.05	0.53	0.48	0.54	0.55	0.28	0.32
15	0.950	0.675	0.095	0.710	0.665	0.665	0.535	0.650	0.655	1.00	0.61	0.15	0.70	0.62	0.60	0.58	0.61	0.59
16	0.950	0.770	0.125	0.845	0.820	0.745	0.710	0.760	0.695	0.97	0.83	0.06	0.79	0.65	0.67	0.74	0.77	0.65

Table 4: Average utility values collected using TTO for session 1 and 2.

Overall average									
Participant	Healthy	Deaf	Worst	A SiS	A SiS+B	AT SiS	AT SiS+B	T SiS	T SiS+B
1	1.00	0.23	0.00	0.41	0.28	0.34	0.25	0.30	0.25
2	1.00	0.77	0.01	0.77	0.74	0.70	0.68	0.71	0.69
3	0.97	0.71	0.01	0.74	0.63	0.66	0.59	0.64	0.61
4	0.92	0.64	0.20	0.74	0.70	0.71	0.69	0.70	0.70
5	0.99	0.65	0.01	0.96	0.80	0.79	0.72	0.85	0.68
6	0.91	0.61	0.01	0.53	0.21	0.30	0.18	0.35	0.21
7	0.96	0.65	0.10	0.78	0.73	0.73	0.73	0.75	0.75
8	1.00	0.72	0.11	0.64	0.61	0.62	0.58	0.61	0.58
9	1.00	0.75	0.00	0.73	0.65	0.67	0.67	0.65	0.62
10	0.98	0.75	0.00	0.90	0.87	0.89	0.89	0.89	0.88
11	0.98	0.55	0.01	0.87	0.81	0.86	0.76	0.74	0.68
12	0.95	0.77	0.05	0.67	0.65	0.61	0.60	0.61	0.63
13	0.96	0.70	0.11	0.90	0.85	0.84	0.85	0.83	0.80
14	0.65	0.46	0.11	0.51	0.49	0.42	0.50	0.40	0.36
15	0.97	0.64	0.12	0.71	0.64	0.63	0.56	0.63	0.62
16	0.96	0.80	0.09	0.82	0.74	0.71	0.72	0.77	0.67
Average utility	**0.95**	**0.65**	**0.06**	**0.73**	**0.65**	**0.66**	**0.62**	**0.65**	**0.61**

Table 5: Average utility values collected using TTO across all repeats and sessions

The fourth aim of the study was to address whether valuations obtained using A alone are susceptible to influence from information gained from other sources; i.e. from the text-based descriptions provided within this experiment. A paired samples t-test was performed comparing the first presentation of the A alone in Block 1, and the second presentation of A alone in Block 3, by which time the participants had read the related text description several times. For SiS, there was no significant difference between the initial presentation of A alone SiS (M=0.73, SD=0.21) and the second presentation of A alone SiS (M=0.65, SD=0.23) t(15)= 1.24 p=0.24. For SiS+B, there was a significant difference between the initial presentation of A alone SiS+B (M=0.68, SD=0.18) and the second presentation of A alone SiS (M=0.56, SD=0.24) t(15)=2.34 , p=0.03. ICC values were also calculated to assess the repeatability of the utility values from the first presentation of the A alone condition and the second presentation of A alone for both A SiS and A SiS+B. For A SiS the within-session (Block 1 vs Block 3) single measure ICC was .35 with a 95% confidence interval from -0.13 to 0.71, for A SiS+B the single measure ICC was .39 with a 95% confidence interval from -0.05 to 0.72.

The fifth aim of the study was to investigate whether valuations collected using A alone or AT are stable compared to valuations collected using T alone. ICC values were calculated for repeated valuations of the same health state description within session 1 and 2, and also across the two sessions (Table 6). ICC values varied widely but were generally lower for within-session reliability in session 1, and when comparing the second exposure of a condition in session 1 to the first exposure in session 2. ICC values between two sessions that consistently exceeded the 0.75 threshold and could be considered stable were only achieved by comparing the average values in session 1 to the average values in session 2.

		Session 1 (exposure 1 and 2)	Session 2 (exposure 3 and 4)	Between sessions (exposure 2 and exposure 3)	Overall across sessions (average of session exposures)
		ICC Single Measures			ICC Average Measures
A SiS	TTO	0.35 (-0.13 - 0.71)	0.93 (0.80 - 0.98)	0.52 (0.02 - 0.82)	0.78 (0.32 - 0.93)
	VAS	0.64 (0.22-0.86)	0.93 (0.81 – 0.97)	0.48 (-0.02 – 0.78)	0.73 (0.10 - 0.83)
A SiS+B	TTO	0.39 (-0.05 – 0.72)	0.94 (0.83 - 0.98)	0.69 (-0.05 - 0.91)	0.92 (0.74 - 0.98)
	VAS	0.45 (0.003-0.76)	0.94 (0.84 – 0.98)	0.82 (0.56 – 0.93)	0.92 (0.79 – 0.97)
T SiS	TTO	0.95 (0.85 – 0.98)	0.99 (0.97 - 0.99)	0.66 (0.19 - 0.89)	0.84 (0.46 - 0.95)
	VAS	0.91 (0.76 – 0.97)	0.98 (0.96 – 0.99)	0.62 (0.22 – 0.85)	0.78 (0.40 – 0.92)
T SiS+B	TTO	0.82 (0.55 – 0.93)	0.97 (0.91 - 0.99)	0.67 (0.04 - 0.90)	0.87 (0.59 – 0.96)
	VAS	0.93 (0.82- 0.98)	0.96 (0.89 – 0.99)	0.89 (0.71 – 0.96)	0.92 (0.78 – 0.97)
AT SiS	TTO	0.86 (0.65 – 0.95)	0.99 (0.95 - 0.99)	0.46 (-0.11 - 0.81)	0.88 (0.56 - 0.97)
	VAS	0.24 (-0.28- 0.65)	0.99 (0.97-0.99)	0.76 (0.49 – 0.91)	0.79 (0.40 – 0.93)
AT SiS+B	TTO	0.65 (0.27 – 0.86)	0.97 (0.88 - 0.99)	0.62 (0.06 - 0.87)	0.90 (0.70 - 0.97)
	VAS	0.78 (0.49 – 0.92)	0.92 (0.78 – 0.97)	0.75 (0.43 – 0.91)	0.89 (0.68 – 0.96)

Table 6: ICC values for A SiS scores session 1, 2 and between sessions, A SiS+B scores session 1, 2 and between sessions, T SiS scores session 1, 2 and between sessions, T SiS+B scores session 1, 2 and between sessions. Values written in red show ICC values that did not meet the criteria for good reliability as per guidance by Koo &Li (2016). Values in brackets are 95% confidence intervals).

The final aim of Experiment 1 was to investigate whether valuations collected using one presentation method correlated with valuations collected using other presentation methods. For example, do valuations collected using ASH alone correlate to valuations collected using T alone and do valuations collected using ASH and text in combination correlate to valuations collected using T alone? This was investigated by looking at the correlation between values obtained using the various presentation modalities (Table 7). Correlation values were almost exclusively high or very high, suggesting that there was some consistency in the information being conveyed through acoustic simulations and text-based descriptions.

		Session 1	**Session 2**	**Overall average**
A vs T				
SiS	TTO	r 0.89**	r 0.88**	r 0.97**
	VAS	r 0.82**	r 0.85**	r 0.97**
SiS+B	TTO	r 0.78**	r 0.98**	r 0.97**
	VAS	r 0.88**	r 0.99**	r 0.97**
AT vs T				
SiS	TTO	r 0.91**	r 0.94**	r 0.97**
	VAS	r 0.54*	r 0.97**	r 0.91**
SiS+B	TTO	r 0.95**	r 0.96**	r 0.96**
	VAS	r 0.91**	r 0.99**	r 0.97**

*Table 7: Pearson's r values for average SiS scores for A and T in session 1, 2 and between sessions and values for average SiS+B scores for A and T in session 1, 2 and between session. * Correlation significant to 0.05 level. ** Correlation significant to the 0.01 level.*

3.4.4 Discussion

There were several questions that this initial experiment was designed to answer, with some of these answered clearly and others not. It appears that A alone descriptions of hearing-related health states can communicate the fact that people with hearing loss are seen as having less than perfect health as the health utility results collected using TTO showed a significant difference between the healthy reference state and A alone SiS and the healthy reference state and A alone SiS+B. This means that ASH can communicate some aspects of ill health as participants did view health states presented using A alone as less than perfect, as established by published assessments of the effects of hearing loss on quality of life (Summerfield et al., 2010, Summerfield et al., 2002). However, simulations may not convey useful additional information that is not already conveyed via traditional descriptive methods as no significant difference was observed between utility values when health states were described using AT vs T. This observation could indicate a limitation of using ASH as a method of communicating health or it could also be that the methodology of applying the simulations to the health valuation process did not effectively communicate the intended difficulties with hearing.

To evaluate the susceptibility of ASH to influence from other sources of information (i.e. the text-based descriptions presented in this experiment) the reliability of the valuations collected in this experiment was assessed. When comparing the first evaluation of the SSD health state collected using ASH alone, when the participant's decisions have not yet been influenced by any information in the text, and the second evaluation of the same health state there was evidence of insufficient reliability. All ICC values and their confidence intervals fell below the level of good reliability identified by Koo and Li (2016). This failure to achieve sufficient reliability suggests that the participants could have been influenced by the text descriptions presented within session 1 of this experiment. However, it cannot entirely be attributed to the influence of the text-based descriptions on the valuations participants gave, as it could also be attributed to the lack of familiarity the participants had with the valuation process, the ASH themselves, the hearing condition or the web-based presentation method. It is also worth noting that all other presentation methods also showed varying levels of reliability.

The ICC values for the health valuations collected using the text-based description, the 'gold standard' health state presentation also displayed mixed reliability. Valuations in the T SiS condition were reliable within session 1 and session 2, but were unreliable between sessions (all lower bounds of confidence intervals for T SiS go below 0.75). It could be expected that the text-based descriptions would be highly reliable, given that one could assume that the meaning is conveyed clearly and consistently to all participants. However, the lack of clear reliability in the data suggests that there may be confounding influences on the valuations collected. Generally, reliability data for A SiS, A SiS+B, T SiS and T SiS+B across and within sessions was suboptimal. There was also poor reliability between the testing sessions with only 4 conditions achieving good levels of reliability and no lower bound of any ICC value for between session valuations reaching the level of good reliability. This could mean that health state valuations, regardless of presentation method and utility collection method, may be susceptible to influence from unknown external sources which occurred between the two separate testing sessions. It could also be the case that due to the novel nature of the simulations participants need to see more examples and familiarise themselves with the simulations and the process of valuing them before valuations are collected.

As discussed previously there is a large amount of variance within the ICC data for A presentations and T presentations. The picture was similar for AT presentations. Session 1 AT SiS collected using TTO shows good reliability however, it is worth noting that the lower bound of the confidence intervals was not above 0.75, and AT SiS collected using VAS was also not reliable. Within session 2, AT SiS utility values were reliable for values collected using both TTO and VAS. Between sessions 1 and 2, AT SiS collected using TTO were unreliable and AT SiS collected using VAS were at the limit of good reliability with the lower bound of the confidence interval not within the acceptable parameters. Valuations collected across both sessions using AT to present SiS were reliable when collected using both TTO and VAS, however again it is worth noting that the lower bounds of the confidence interval for TTO and VAS related ICC values do not surpass 0.75, the acceptable parameters for reliability. The lack of clear reliability for valuations collected using T alone is a concern, as text-based descriptions of health states are considered the ‘gold standard’ method for describing complex and rare health states.

The correlation between presentation methods was evaluated in order to access whether valuations collected using A alone were related to valuations collected using T and if valuations collected using AT were related to valuations collected using T. The significant positive correlations between SiS and SiS+B listening situations across all presentation conditions and testing sessions indicate that there is a positive relationship between variables collected using A and T and AT and T. This seems to imply that the listening situations portrayed by ASH, a novel presentation method, in the A and AT conditions do portray the listening difficulties of individuals living with SSD as the validated text-based descriptions of SSD developed by Lucas and Kitterick (2014) do. However, these results must be interpreted cautiously until the effectiveness of the simulations to convey adequately the listening difficulties experienced by those with SSD has been assessed, which was the aim of Experiment 2.

3.5 Experiment 2

3.5.1 Introduction

Text based descriptions must be validated to ensure the language they use adequately conveys the difficulties of having a certain health condition. Similarly, ASH need to be examined to ensure that they are adequately describing the relevant difficulties of having a specific form, or degree, of hearing loss. Experiment 2 used a behavioural task to examine whether the simulations used in Experiment 1 not only represented the difficulties encountered by those with SSD as stated in the focus groups ran by previous research (Lucas & Kitterick 2018) but also recreated those difficulties for participants. The behavioural task also aimed to investigate whether the simulations adequately demonstrate that monaural listening, listening to sound when it is only presented to one ear, is more difficult than binaural listening, listening to sound when it is presented to two ears simultaneously, and that SiS is easier to follow than SiS+B, in both monaural and binaural listening situations. If the ASH are not more difficult in both the SiS+B setting and in the monaural situations, then the simulations are not adequately recreating the hearing-related health states described by patients with SSD; i.e. they do not have face validity. The behavioural task also aimed to ensure that there are no ceiling or floor effects within the simulations, meaning that there is not a simulation in which it is easier to decipher the talkers than would be expected and vice versa.

3.5.1.1 *Aims and research questions*

Experiment two aims to address the following research questions:

1. Do the simulations adequately demonstrate that binaural listening is more difficult than monaural listening?
2. Do the simulations adequately demonstrate that SiS is easier to follow than SiS+B in both monaural and binaural listening situations?

3.5.2 Methods

3.5.2.1 *Participants*

This experiment was conducted in accordance with the permissions granted by the NHS Grampian Research Ethics Committee and the Sponsor (University of Nottingham) Written informed consent was obtained for all participants by the Principal Investigator. Sixteen normal hearing participants (6 males, 10 females) with no expert knowledge of hearing research or audiology took part in the study. Nine of the participants had previously taken part in Experiment 1, as Experiment 2 occurred in tandem with Experiment 1. Participants had an age range of 18 to 52 years old (mean age 27.8 years) and all participants had pure-tone average threshold ≤ 30 dB HL [Hearing Level] in both ears between 0.5 and 4 kHz and did not have a current ear infection or any other condition that would prohibit a hearing assessment taking place prior to the first testing session. Participants were recruited from the National Institute for Health Research Nottingham Biomedical Research Centre (NIHR Nottingham BRC) participant database, and from posters placed in student areas on the University of Nottingham University park campus. Participants were financially compensated for their time at a rate of £5 for up to one hour of participation and reimbursed for any incurred travel expenses. The experiment lasted approximately 45 minutes.

3.5.2.2 *Design*

The study was a 2x2 factorial design. Four conditions were defined from the combination of the two listening configurations (monaural and binaural) and the two listening environments (SiS and SiS+B). Each participant took part in two sessions. In each session, 25 example simulations were presented in each condition, resulting in a total of 100 simulations in each session. Each participant was asked to attend to the talker in front of them (talker A, see figure 1 and 2). The task was to write down the last sentence that talker A said. The simulations presented in Experiment 2 were the same simulations as those used in Experiment 1 as detailed previously.

3.5.2.3 *Counterbalancing*

To control for order effects, the conditions in Experiment 2 were counterbalanced using a Latin Square design (Graziano and Raulin, 2014). Order effects refer to the differences in participants' responses that result from the order in which experimental materials are presented. Order effects are of special concern in within-subject designs when the same participants are in all conditions and responses are to be compared between conditions. A 4x4 Latin Square was used to counterbalance the conditions presented to each participant. The 4x4 Latin Square sequence was repeated for every fourth participant to ensure that the specific list of sentences used for talker A (the target talker, see *Procedure* below) were not repeated within each participant's testing session. Complete counterbalancing was used so that all possible orders of conditions occurred, that all conditions were presented an equal number of times in each position in the testing order, and each condition preceded and followed each other condition an equal number of times.

3.5.2.4 *Procedure*

Participants gave informed consent to participate in the study and to a hearing screening being carried out by the researcher. Once the consent was obtained a screening audiogram was obtained to ensure participants had a PTA threshold ≤ 30 dB HL [Hearing Level] from 0.5 to 4 kHz(Audiology, 2011). Simulations were calibrated using the same procedure as used in experiment 1 (3.4.2.1). Simulations were presented at 64.9 dB in the left ear, normal hearing ear, and 0 dB in the ear with simulated profound deafness. A level of 65 dB was chosen as this is the sound level for normal speech, as the simulations were presented at 64.9 dB the level was within the 1 dB tolerance level set by the researcher.

Participants were seated at a computer which was situated in a double-walled sound-attenuated room and asked to put on a pair of Beyerdynamic DT 770 Pro Studio closed-back headphones so that they comfortably enclosed their pinnae. Participants were then asked to listen to the SiS and SiS+B ASHs and write down the third sentence spoken by the talker at position A, described as speaking from directly in front of them. This talker was chosen as the target because it would only be possible to determine its location in the binaural listening condition; i.e. it would maximise the difficulty that participants would experience in identifying the correct talker to listen to in the monaural simulations that represented the health state of having SSD. Once the testing session was completed, participants were invited back for a second testing session to repeat this procedure a second time.

3.5.3 Analysis

Each target sentence contained 5 key words. For example, in the sentence “The birch canoe slid on the smooth planks” the key words were birch, canoe, slid, smooth, and planks. The ability of the participants to hear sentences presented from the target talker was summarized by calculating the proportion of words in the sentences they could correctly identify immediately after presentation.

3.5.4 Results

The accuracy with which participants could report key words in the target sentence across the four listening conditions is summarised in Table 8 for both session one and session two. There appeared to be an effect of ears as the poorest performance was observed in a monaural condition (SiS+B in session one: 72.2%) and the highest performance was observed in a binaural condition (SiS in session two: 96.2%).

	Binaural			Monaural		
	Session 1	*Session 2*	*Average*	*Session 1*	*Session 2*	*Average*

SiS	84.9% (21.1)	96.2% (5.4)	90.5% (11.7)	87.7% (14.3)	95.3% (7.1)	91.5% (10.1)
SiS+B	87.1% (9.0)	91.7% (6.7)	89.4% (7.3)	72.2% (16.7)	83.0% (10.9)	77.6% (13.1)

Table 8: Percentage of key words identified across conditions

The data were subjected to a repeated measures ANOVA with factors of session, condition (quiet vs noise), and ears (binaural vs monaural). The main effects of condition ($F(1,15)=14.4$, $p<.01$), of session ($F(1,15)=23.1$, $p<.001$), and ears ($F(1,15)= 4.7$, $p<.05$) were significant. Performance was poorer in noise than in quiet (mean difference 7.5%), poorer in session 1 than in session 2 (mean difference 8.6%), and poorer with one ear than with two (mean difference 5.4%). There was a significant interaction between condition and ears ($F(1,15)= 14.3$, $p<.01$). Post-hoc comparisons revealed that there was a significant effect of ears, but only in the noise condition (mean difference 11.8%, $t(15) = 6.0$, $p<.001$) and not in the quiet condition (mean difference 1.0%, $t(15)=0.3$, $p > .05$).

3.5.5 Discussion

The results of the listening task appear to confirm that there is an effect of condition, session, and ears. However, contrary to expectations, performance was generally high and exceeded 70% in all conditions across both sessions, even when the simulations were designed to ensure that participants had only one functioning ear (monaural conditions). This result could have been due to the participants using level cues within the simulations to attend to the correct talker; i.e. rather than trying to work out which sentence was being said by the target talker the participants were paying attention to the loudest talker. To eliminate this possibility, it would be necessary to rove the levels at which the talkers speak the sentences, but this would come at the cost of the simulations losing some of their ecological validity; i.e. their realism. Initially the aim of this thesis was to set up simulations of real world environments in a manner which is as life like as possible, but attempting to do so without other sensory data present may mean that the simulations need to be more contrived so that they effectively convey the difficulties of having a hearing loss.

The lack of a statistically significant difference between monaural and binaural performance in the SiS condition could have also been attributed to the high levels of performance, perhaps resulting from these loudness cues, which could have led to at or near ceiling effects that made it more difficult to detect differences between SiS conditions. It could also have been the case that the background noise used in the SiS+B condition was set at too low of a level and needed to be higher (and more reflective of signal to noise [SNR] encountered in everyday life) to ensure that the simulations adequately communicate the listening difficulties those with SSD have; i.e. that listening to speech in background noise is more difficult than listening to speech in quiet.

There could also have been an effect of experimental condition as the participants became more practised at the listening task and consequently scored higher in session 2 as they became more adept at using possible level cues to identify the target talker.

3.6 General Discussion

The results of pilot 1 indicated that using ASH could communicate the fact that people with hearing loss are seen as having less than perfect health could be feasible, although the effect of presentation modality was not clear. Experiment 1 indicated that ASH presented alone can communicate the fact that people with hearing loss are seen as having less than perfect health. However, the valuations obtained suggest that currently the simulations may not convey useful additional information that is not already conveyed via traditional descriptive methods. Failure to achieve acceptable levels of reliability suggests that participants could have been influenced by the text descriptions presented within the experiment. The correlation between presentation methods was evaluated. Experiment 2 reported the results of a listening task and appeared to confirm that there is an effect of listening condition (SiS vs SiS+B), study session, and listening modality (monaural vs binaural).

As ASH are a novel and yet untested methodology for presenting health states there were several fundamental questions that surrounded ASH’s suitability as a health description tool. The research carried out in this chapter needed to address the concerns surrounding feasibility before ASH could be developed further and utilised as a health state presentation method. The results of pilot 1 indicated that a basic simulation of a hearing-related health state, in the case of this research project SSD, was theoretically feasible. The difference between SiS and SiS+B utility values in pilot 1 also indicated that the ASH could also be a successful in communicating not just the difficulties caused by a hearing loss, but also the different levels of difficulty that can be experienced by someone living with hearing loss to participants with healthy hearing.

Pilot 1 had a limited scope and it was important to address in experiment 1 not only whether ASH can convey health states with a hearing loss, but also healthy hearing. It was important for experiment 1 to address whether an ASH of healthy hearing are as feasible, useful and stable as ASH of SiS and SiS+B health states are and that the alterations to the simulations had not hindered the ability to simulate SSD related health states as observed in pilot 1. Although the findings from experiment 1 appeared to show that the revised ASH do still convey listening difficulties and that ASH can be used to present hearing-related health states in health valuation exercises, the results also highlighted a lack of consistent reliability within the A alone data sets and within data sets collected across testing sessions.

The valuation process did not take a long time to complete using the web-based platform, so it may be more practicable to collect multiple repeated utility values within one session to maximise their reliability. This would also reduce the opportunity for any confounding external factors present between testing sessions to influence the utility values collected within the experiment and could increase the overall reliability of the utility value data.

The text descriptions of the hearing-related health states were reliable. However, the large spread in the confidence intervals for text presentation reliability values in experiment 1 was unexpected. This result could also have been due to the way that the presentations of health states were displayed to participants. It is unusual to present health valuation exercises without a researcher introducing the material to the individual rating the health state or being available to guide them through the valuation process in some way. This lack of input from a researcher, the passive way that health state descriptions were displayed and using a novel web-based platform may have prevented participants from fully engaging with and/or understanding the simulations and health state descriptions. Participants passively engage with the descriptions and simulations, meaning that they are presented on screen and participants are left to read the descriptions or listen to the simulations at their own leisure. Further research into this topic is carried out in experiment 5 (see section 5.4) to investigate the effects of more active engagement, such as having to answer questions about the health state descriptions presented and any improvement in the reliability of values collected.

It may also be that the novel description method and the presentation of the health state descriptions are alien to the participants and they simply need more presentations of the health states in order to familiarise themselves with the ASH, the presentation methods, and the valuation process.

Experiment 2 answered questions surrounding the ability of ASH to convey the specific listening difficulties intended. The results observed in Experiment 2 show that although the revised simulations were better at simulating hearing difficulties than the initial versions, it is still the case that the listening difficulties portrayed by the ASH did not fully align with the listening difficulties reported by SSD patients in the development of Lucas and Kitterick's (2014) validated text descriptions of living with SSD. The simulations appear to only partially recreate the listening difficulties typically experienced by those living with SSD, such as difficulties identifying listeners and listening to speak in noisy situations when other talkers are present. There were significant differences in the number of key words noted by participants between SiS and SiS+B conditions within experiment 2, however, the performance levels were overall quite high. These higher-than-expected performance levels suggest that the simulations could be conveying some of the difficulties but not all of the desired difficulties, that the ASH are not performing to their full potential and would benefit from further small alterations. The simulations needed to be altered to make it harder for those with normal hearing to discern the individual speakers voices but the ASH still need to maintain a level of realism and ecological validity so that they can be relatable and tangible to a participant rating the healthy state.

To conclude, ASH appear feasible as a health description tool and can convey different levels of difficulty, but still need development. Single testing sessions are more practical and lead to less variability within the utility data. Further experiments within this body of work will need to investigate the effects of passive and active engagement with the stimulus on health utility values, and their reliability, as well as how the frequency and volume of presentations influence health utility values.

4 Further development of acoustic simulations

4.1 Introduction

Experiments 1 and 2 indicated that the use of ASH to describe hearing-related health states is feasible with participants able to perform both a TTO and a VAS task when presented with health states described by ASH alone, text alone, or a combination of both. Results from Experiment 2 suggest that the simulations ability to recreate one of the main difficulties of monaural listening, sound localisation, need to be improved in order to adequately recreate what it’s like to hear with SSD. As discussed previously, the short falls of the revised simulations, version 2, could be due to an unexpected level cue (the target talker being consistently the loudest), an unexpected timing cue (the target talker consistently starting to speak at the same time), or in the case of the SiS+B simulations the signal-to-noise ratio (SNR) of the background cafeteria noise not being sufficiently poor to reflect real-world environments. It is vital that these problems are addressed as the reliability of simulations, and their ability to consistently recreate the listening difficulties of SSD, could affect the reliability of valuations obtained using them.

The aim of work detailed in this chapter was to improve the simulations and their ability to communicate the main listening difficulties of SSD. The objective of the work detailed in this chapter was to modify and improve the simulations by removing any level cues, investigate whether these changes influence behavioural performance when listening to the simulations, and to confirm whether the SNR chosen is appropriate.

4.2 Version 3 simulations - Alterations made to version 2 simulations

No changes were made to the simulation environment or simulation set up; all simulations were again situated in the cafeteria environment. As with version 1 and 2 of the simulations, the sentences presented from the A position were sentences from the IEEE sentence corpus (1969) at 2, and 10 seconds approximately. During this time there is a back and forth conversation between speakers B and speakers C. The sentences presented from speaker B and C position were sentences from the Bamford-Kowal-Bench (BKB) sentence corpus (Bench et al., 1979).

To reduce any level cues that could induce participants into systematically listening to the loudest talker roving was used. This brought the level of all talkers to similar levels. Roving referred to the level of each talker, which was varied over a wide decibel range but that the average presentation level was similar for each speaker location, regardless of where they were located. Thus, within any ASH presentation, if a participant had adopted a strategy of listening to the loudest talker, as a coping strategy in order to overcome the difficulties associated with an inability to localise sounds, then they were equally likely to listen to the talker at position A, B or C (see chapter 3, figure 1)

4.3 Pilot 2: Behavioural validation of simulation changes (floor and ceiling effects)

4.3.1 Introduction

It was necessary to evaluate the changes made to the simulations in order to investigate whether roving the levels of the talkers in addition to the temporal jitter that was added in the previous revision, had together removed any predictability and potential cues as to the identity of the target talker. This pilot experiment explored how well these changes to the simulations had eliminated any possible level and timing ques and if floor and ceiling effects were still present.

4.3.2 Methods

4.3.2.1 Participants

This experiment was conducted in accordance with the permissions granted by the NHS Grampian Research Ethics Committee and the Sponsor (University of Nottingham) Written informed consent was obtained for all participants by the Principal Investigator. All data were anonymised and stored in accordance with General Data Protection Regulations.

Four researchers based at Nottingham BRC participated in the pilot experiment. Demographic data were not collected. The sample size was chosen based on convenience to address differences in performance on a listening task using the revised simulations and was not chosen to achieve sufficient statistical power for a specific statistical comparison. Participants were age 18 years or older, had self-reported normal hearing, and were willing and able to complete a short experiment involving a listening task based on the new simulations. Participants were not financially compensated for completing the study, the pilot experiment lasted approximately 20 minutes.

4.3.2.2 Design

Two conditions were defined from the two listening configurations (monaural and binaural). Each participant took part in one session. In the session, 25 example simulations were presented in the SiS listening configuration. Only the SiS listening situation was presented as this was the easier listening situation to understand the speech presented to them. If participants could not determine what the target sentence was in a quiet environment, then it would follow logically that they would also be unable to determine what the target sentence was if background noise was present. Each participant was asked to attend to the talker in front of them (talker A, see 3 3.2.2). As in Experiment 2 the task was to write down a target sentence, which was the last sentence that talker A said.

4.3.2.3 Procedure

As with pilot 1, participants were seated at a computer which was situated in a double-walled sound-attenuated room and asked to put on a pair of closed-back Beyerdynamic DT 770 Pro Studio headphones so that they comfortably enclosed their pinnae. Participants were then asked to complete a behavioural task in which they were asked to listen to the SiS ASH presented monaurally and binaurally in a random order, and write down the third sentence spoken by the talker at position A.

4.3.2.4 Analysis

Each target sentence contained 5 key words. For example, in the sentence “Note closely the size of the fuel tank” the key words were ‘note’, ‘closely’, ‘size’, ‘fuel’, and ‘tank’. As in Experiment 2, the ability of the participants to hear sentences presented from the target talker was summarized by calculating the proportion of words in the sentences they could correctly identify immediately after presentation and loose key word scoring was used (see section 3.5.2.2).

4.3.3 Results

The average listening performance did reach ceiling level for three listeners but only for binaural listening, as expected. Average key word identification for the monaural listening level was approximately 70 percentage points lower than that of binaural listening.

	Percentage of key words correctly identified				
	P1	P2	P3	P4	Overall average
Monaural	16%	44.8%	15.2%	38.4%	28.6%
Binaural	95.2%	100%	100%	100%	98.8%

Table 9: Average performance for monaural listening was lower than binaural listening. P1 = participant 1, etc.

4.3.4 Discussion

It appears that the changes to the simulations improved their ability to convey an increased difficulty when localising sounds when listening monaurally as key word identification in monaural listening was lower than levels seen in Experiment 2 (3.5.3). Although the average listening performance did reach ceiling levels for three listeners within the binaural listening conditions it is worth noting that these are experienced listeners, which could have influenced the amount of key words recorded. Furthermore, the simulation situation used was SiS which did not include any background noise and consequently it would have been trivial to discern and report the target sentences. There is a large amount of variability within the data collected, but previous research has suggested that variability in data is to be expected. Welsh (2004) found substantial variability between the listening abilities of participants completing a similar listening task.

Although there are ceiling effects present in some cases, and high levels of variability present in the data, the changes made to the simulations suggests that the simulations appear far more suitable and demonstrate monaural listening difficulties much more effectively than the previous simulations.

4.4 Pilot 3: Background SNR & Behavioural validation of simulation changes

4.4.1 Introduction

The SNR ratio chosen for the initial simulations of the SiS+B environment was an SNR of +5 dB; i.e. the speech was presented at a presentation level that was 5 dB louder than the background noise. This SNR was chosen during the initial development process as it was believed that this SNR would still allow speech to be heard over the background cafeteria noise but with some difficulty. This SNR choice was supported by an article by Smeds et al. (2015), who carried out real world recordings to measure the average SNR encountered by hearing aid users. An SNR of +5dB was the most common encountered in day-to-day scenarios. This pilot study explored whether the SNR chosen in experiments 1 and 2 was appropriate, i.e. the chosen SNR was most effective at exposing a difference between binaural and monaural listening conditions and if the originally chosen SNR of +5 dB had affected the high levels of performance seen in experiment 2.

4.4.2 Methods

4.4.2.1 Participants

This experiment was conducted in accordance with the permissions granted by the NHS Grampian Research Ethics Committee and the Sponsor (University of Nottingham) Written informed consent was obtained for all participants by the Principal Investigator. All data were anonymised and stored in accordance with General Data Protection Regulations.

Five researchers based at Nottingham BRC participated in the pilot experiment. Demographic data were not collected. The sample size was chosen based on convenience to address differences in performance on a listening task at different SNR and was not chosen to achieve sufficient statistical power for a specific statistical comparison. Participants were age 18 years or older, had self-reported normal hearing, and were willing and able to complete a short experiment involving a listening task based on the latest simulations. Participants were not financially compensated for completing the study, the pilot experiment lasted approximately 20 minutes.

4.4.2.2 Design

The study was a 2x3 factorial design. Six conditions were defined from the combination of the two listening configurations (monaural and binaural) and the three SNR for the listening environments (+3 dB, +5dB, +7dB). Each participant took part in one session in which 30 example simulations were presented in each listening configuration, with 10 presented at each possible SNR, resulting in a total of 60 simulations in the session. Each participant was asked to attend to the talker in front of them (Figure 1, section 3.2.2.). The task was to write down a target sentence, which again was the last sentence that talker A said.

4.4.2.3 Procedure

Participants were seated at a computer, which was situated in a double-walled sound-attenuated room and asked to put on a pair of closed-back Beyerdynamic DT 770 Pro Studio headphones so that they comfortably enclosed their pinnae. Participants were then asked to complete a behavioural task in which they were asked to listen to the SiS+B AS presented at +3 dB, +5 dB and +7 dB and write down the third sentence spoken by the talker at position A.

4.4.2.4 Analysis

Each target sentence contained 5 key words. For example, in the sentence “A lot of sugar makes sweet tea” the key words were ‘lot’, ‘sugar’, ‘makes, ‘sweet’, and ‘tea’. As in Experiment 2, the ability of the participants to hear sentences presented from the target talker was summarized by calculating the proportion of words in the sentences they could correctly identify immediately after presentation and loose key word scoring was used (see section 3.5.2.2).

4.4.3 Results

The average percentages of correct key words identified are presented in Table 10. On average and as would be expected, an SNR of +3dB produced the lowest proportions of key words identified for both monaural and binaural listening, and a +7 dB SNR resulted in the highest percentages of correctly identified key words. An SNR of +5 dB produced mid-range levels of correctly identified key words for monaural and binaural listening, but importantly demonstrated the largest difference in accuracy between monaural and binaural (26.5%)

Listening condition	SNR	Percentage of key words correctly identified
Monaural	3	36%
	5	60%
	7	74%
Binaural	3	52%
	5	86.5%
	7	92%

Table 10: Percentage of key words identified in each SNR for binaural and monaural presentations

4.4.4 Discussion

It is important that the SNR chosen for the ASH makes it difficult for participants to perform a listening task as those with SSD report that they do struggle with listening monaurally in noisy situations, but those same individuals also report that it is not absolutely impossible to understand speech (Lucas et al., 2018, Welsh et al., 2004).

All SNR levels produced performance levels that were approximately 20% lower when listening monaurally than when listening binaurally. An SNR of +3dB could be chosen for simulations as it produced performance levels that were very poor when listening monaurally, but it also produced performance levels that were quite poor for binaural listening. It is worth noting that the participants in this experiment are all experienced listeners and if these participants were performing at around 50% binaurally it is logical to surmise that inexperienced listeners may perform worse, risking the performance at this task hitting floor level. An SNR of +5 dB appears most suitable to use in future experiments as it ensured that there were no ceilings or floor effects present when participants were asked to identify sentences. An SNR of +5dB made listening to speech in a noisy situation difficult but not impossible, with binaural hearing and more difficult than monaural hearing. Also, if inexperienced listeners were to perform worse than these experienced listeners then with an SNR of +5dB they would not be likely to be at floor. This conclusion is further supported by Smeds et al (2015) who found that the average SNR for speech in noisy environments was approximately +5 dB in realistic sound environments encountered by hearing aid users.

4.5 Experiment 3: Full behavioural validation

4.5.1 Introduction

As discussed in Experiment 2 (Section 3.5), acoustic simulations of hearing loss need to be examined to ensure that not only are they adequately describing the relevant difficulties of having a hearing loss but also recreating those difficulties accurately. Experiment 3 used the same behavioural task as Experiment 2 to examine whether the latest revisions to the simulations now accurately recreated and represented the difficulties encountered by those with SSD (Welsh et al., 2004, Borton et al., 2010, Lucas et al., 2018, Vermeire and Van de Heyning, 2009, Wie et al., 2010, Heffernan et al., 2016). The behavioural task aimed to investigate whether the simulations adequately demonstrate that binaural listening is more difficult than monaural listening and that SiS is easier to follow than SiS+B, in both monaural and binaural listening situations. As with Experiment 2, this experiment also aimed to confirm that there were no ceiling or floor effects present.

4.5.1.1 Summary of research questions

Experiment 3 aims to address the following research question:

3. Do the simulations adequately demonstrate that monoaural listening is more difficult than binaural listening?
4. Are there floor and ceiling effects present?

4.5.2 Methods

4.5.2.1 Participants

Sixteen normal hearing participants (2 males, 14 females) with no expert knowledge of hearing research or audiology took part in the study. None of the participants had previously taken part in any previous experiments related to this thesis. Participants had an age range of 21 to 44 years old (mean age 25.81 years) and all participants had pure-tone average threshold ≤ 30 dB HL [Hearing Level] in both ears from 0.5 to 4 kHz and did not have a current ear infection or any other condition that would prohibit a hearing assessment taking place prior to the first testing session. Participants were recruited using posters placed in student areas on the University of Nottingham University park campus. Participants were financially compensated for their time at a rate of £5 for up to one hour of participation and reimbursed for any incurred travel expenses. The experiment lasted approximately 45 to 60 minutes.

4.5.2.2 Design

The study was a 2x2 factorial design. Four conditions were defined from the combination of the two listening configurations (monaural and binaural) and the two listening environments (SiS and SiS+B). Each participant took part in one session. In the session, 25 example simulations were presented in each condition, resulting in a total of 100 simulations in each session. Each participant was asked to attend to the talker in front of them (talker A, see figure 1). The task was to write down the last sentence that talker A said. The simulations presented in Experiment 3 were the same simulations as those used in pilots 2 and 3 as detailed previously.

4.5.2.3 Power

The experiment was not powered separately, but replicated recruitment figures for experiment 2. Experiment 2 was powered based on data from Pilot 1, a power analysis for a repeated measures ANOVA with two groups was conducted in G*Power (Faul 2007, 2009) to determine a sufficient sample size using an alpha of 0.05, a power of 0.80, and a medium effect size (f = 0.7) (Faul et al., 2009, Faul et al., 2007). Based on these assumptions, the desired sample size was 14, with a target recruitment figure of 16 to account for possible attrition.

4.5.2.4 Counterbalancing

To control for order effects, the conditions in Experiment 3 were counterbalanced using a Latin Square design (Graziano and Raulin, 2014).

A 4x4 Latin Square was used to counterbalance the conditions presented to each participant. The 4x4 Latin Square sequence was repeated for every fourth participant to ensure that the specific list of sentences used for talker A were not repeated within each participant's testing session. Complete counterbalancing was used so that all possible orders of conditions occurred, that all conditions were presented an equal number of times in each position in the testing order, and each condition preceded and followed each other condition an equal number of times.

4.5.2.5 Procedure

Participants gave informed consent to participate in the study and to a hearing screening being carried out by the researcher. Once the consent was obtained a screening audiogram was obtained to ensure participants had normal hearing. Simulations were presented at 65.1 dB, within the 1 dB tolerance level set by the researcher. Participants were seated at a computer, which was situated in a double-walled sound-attenuated room and listened to simulations using Beyerdynamic DT 770 Pro Studio headphones. Participants were then asked to complete a behavioural task in which they were asked to listen to the SIS and SIS+B ASHs and write down the third sentence spoken by the talker at position A. Once the testing session is completed participants were invited back for a second testing session to repeat this procedure a second time.

4.5.2.6 Analysis

Each target sentence contained 5 key words. For example, in the sentence "We talked about the side show at the circus" the key words were 'talked', 'about', 'side', 'show', and 'circus'. The ability of the participants to hear sentences presented from the target talker was summarized by calculating the proportion of words in the sentences they could correctly identify immediately after presentation.

4.5.3 Results

The overall accuracy with which participants could report key words in the target sentence across the four listening conditions is summarised in Figure 14 and average key words identified in the target sentence across the four listening conditions summarised in Table 11. There appeared to be an effect of environment and ears. The poorest performance was observed in a monaural condition (SiS+B 15.4%) and the highest performance was observed in a binaural condition (SiS 64.4%). The data were subjected to a repeated measures ANOVA with factors of environment (SiS vs SiS+B), and ears (binaural vs monaural). The main effects of ears ($F(1,15)= 75.94$, $p<.001$) and environment ($F(1,15)=7.89$, $p<.01$) were significant. Performance was poorer in SiS+B than in SiS (mean difference -6.7%), and poorer with one ear than with two (mean difference 48.4%). There was no significant interaction between environment and ears.

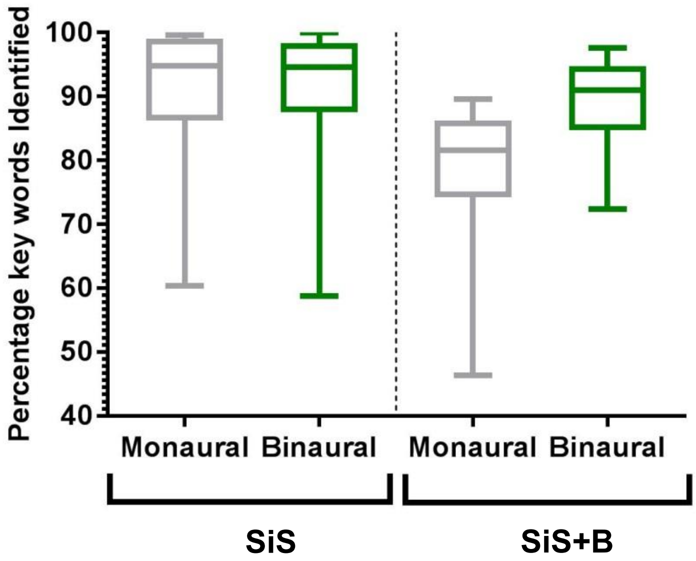


Figure 14: Percentage of keywords identified by participants

	Monaural	Binaural
SiS	*15.4%*	*64.4%*
SiS+B	9.2%	57.1%

Table 11: Mean percentage of key words identified across conditions

4.5.4 Discussion

The results of the listening task appear to confirm that there is an effect of condition (easier in quiet than in noise), and ears (easier with two ears than with one). It is also worth noting that compared to Experiment 2 (section 3.5), performance levels overall were considerably lower, as was desired. These lower listening performance levels can be explained by the changes in the simulations that aimed to remove any level cues present within the simulations to attend to the correct talker, i.e. participants could no longer just pay attention to the loudest talker and had instead to try to work out which sentence was being said by the target talker using localisation cues that were effectively disrupted by presenting the simulations monaurally. Despite the lower performance levels, there is still a considerable difference between monaural and binaural listening performance in the SiS and SiS+B listening environments (~50%).

4.6 General discussion

The improvements made to the simulations appear to have improved their ability to convey an increased difficulty when listening monaurally versus listening binaurally as desired. Consequently, floor and ceiling effects appear to no longer be present with listening performance lower overall and significantly lower monaurally than binaurally. Additionally, the SNR chosen originally for the background noise used in the SiS+B condition has been deemed appropriate and had now been supported by empirical evidence (Pilot 3 section 4.4.3) as well as literature (Smeds et al 2015).

Although key issues with the suitability of the simulations developed for this research has been addressed, many questions remain about the most suitable way to present ASH to individuals rating health states. The foremost question about the use of ASH centres around the reliability of such an unproven method. Now that ASH appear to be a feasible method and can convey the listening difficulties of an individual with hearing loss, it is again important to look at how reliable ASH are when used within the valuation process. If valuations of health states collected using ASH are highly unstable and required multiple presentations in order to obtain reliable utility values then the method may not be practical to use, especially when collecting large numbers of valuations, due to the time constraints involved. Therefore, the reliability of ASH needs to be addressed in detail: how variable are valuations obtained using ASH across individual participants, across groups of participants, for different listening situations?

There are also several complex, intertwined questions surrounding the presentation of ASH within the health valuation process. In the first initial explorations into the use of ASH (pilot 1), participants were provided with simulations of both healthy and diseased health states; that is, they could listen to an ASH presented binaurally and then monaurally to understand the difference between them. However, when text descriptions are given to individuals rating a health state, they are usually given text descriptions of the diseased state. This raises the question of whether it is also appropriate to only present simulations of the hearing-related health state being measured or whether both the healthy example and the diseased state should be presented side by side to convey the changes being described.

Simulations are a novel and alien methodology, so participants listening to a disease simulation may not realise what they are listening to if a healthy simulation is not provided to them. Perhaps ASH should in fact be used in combination with the current gold standard method (text) to give a more complete picture of what it is like to live with SSD. Furthermore, simulations are limited in what they can adequately simulate, just as text descriptions are limited in what they can describe, to individuals rating health states. In valuation studies that use text-based descriptions, one relies on the ability of the individuals rating the health states to conceptualise the information presented to them. Presenting both text and ASH together could have complementary effects on both descriptive mediums. For example, ASH cannot convey emotions (i.e. an individual with SSD may be frustrated by their inability to hear in a crowded room, or exhausted by the effort of listening) whereas a text description can convey this information. Similarly, it can be difficult to convey complex ideas such as localisation of sound using a short text description alone, but using an ASH it is possible to give a concrete example of how an individual with SSD cannot easily discern the location of talkers. In short, using ASH and text together could eliminate the shortcomings of both methodologies and bring valuation levels more in line with utility valuations provided by patients who live with the condition on a daily basis.

Lastly, within the previous experiments participants have been provided with simulations and simply asked to listen to them, as happens when they are provided with the text descriptions in health utility analysis. However, it is therefore not possible to know if this is sufficient to ensure that the participants have played the simulations and paid attention to the listening difficulties conveyed within them. It may be the case that the methodology needs to be adapted to facilitate more interaction with the ASH to ensure that participants have paid attention to the simulations. When describing a health state using a text-based description, the participant engages with the description by reading it, but they can listen passively to audio examples. Therefore, the question that arises is whether it is necessary to ensure that participants engage with the ASH and whether any such engagement would have any effect on the reliability of the valuations collected. The next chapters of this thesis aim to address this, and the other questions posed in this discussion.

5 Refining the methodology of applying acoustic simulations.

5.2 Introduction

Work detailed in Chapters 3 & 4 addressed several initial questions surrounding the suitability of the ASH that have been developed, and the feasibility of using those ASH to describe health states within the economic evaluation process. With these questions answered attention can now shift to other questions regarding how best to use ASH in practice. For example, does a participant's attention need to be guided by a task when ASH are used? Would a participant need an ASH which describes what is expected by 'best health imaginable'? Can ASH be used in conjunction with a current validated vignette? Can an ASH be used to describe a treatment option and can an ASH be presented remotely to facilitate easier large-scale data collection.

As discussed previously in chapters 3 and 4, ASH have not been used to describe health states in an economic evaluation process, so there is no existing guidance on exactly how ASH should be presented. As discussed within chapter 2 Goldstein et al. (1994) investigated the effects of using multimedia to present health state descriptions. Goldstein et al (1994) found that when multimedia was used to present the health states via video clips participants were more familiar with the descriptions and recalled more aspects of the health state descriptions. This presents an interesting question, should it be considered best practice to facilitate certain levels of interaction with ASH to ensure that participants have paid attention to the simulations and the information that they are conveying? Would the inclusion of an active task within this research influence the reliability of health state valuations collected? An economic evaluation exercise could be a lengthy task, especially if in the future use of ASH they are presented in multiples to describe several unique scenarios, such as SiS, SiS+B, sound localisation etc. A lengthy task this may lead to attentional lapses due to boredom (Schooler et al., 2004, Manly et al., 1999) leading to lower engagement with the health state descriptions. Adding a task could mitigate any attentional lapses before they occur. Would the addition of such a task require multiple presentations of the health state to ensure acceptable reliability levels are reached?

During previous experiments participants were asked to play a simulation and rate the health state, which could be viewed as a passive task as there was no directed engagement, however the very act of reading a vignette description of a health state could be seen as actively engaging with the description. It could be argued that by reading the contents of the vignette, processing the described effects of the health state on daily life, and possibly visualising and conceptualising how this could affect the participant's own life, a participant is actively engaging with the vignette. This again raises the question of whether participants need to be prompted to actively engage with an ASH in the similar way. It may also be beneficial for such engagement to be facilitated if the data collection process is carried out via remote means, for example through a website. Using a website to administer the valuation tasks would reduce the burden on researchers during data collection, but it would also result in a lack of control of the research environment and a lack of control over how the participant interacts with both the ASH and the valuation tasks. Remote administration of a health valuation task may benefit from controls such as a listening task to guide attention.

It is important to answer these questions fully before considering the potential further uses for ASH and other situations that may be useful to recreate using ASH. Experiments 4 and 5 were designed to address these presentation and engagement questions, a comparison of the results from experiment 4 and 5 was also performed in order to compare the effects of presenting two ASH (one of 'normal' hearing and one of SSD) on health state valuations and their reliability. Experiment 6 was designed to address further questions surrounding the presentation of ASH alongside text, the experiment was presented as a 'second part' to both experiments 4 and 5; although this experiment was carried out within the same study sessions as experiments 4 and 5 it was a stand-alone study.

5.3 Experiment 4: Is an active or passive task needed to promote engagement with ASH if an ASH of ‘healthy hearing’ is presented?

5.3.1 Introduction

Experiment 4 was designed to answer the questions surrounding whether there will be an effect of environment when participants are not provided with an example of normal hearing. Previous experiments (see chapters 3 and 4) have shown that ASH can portray differing levels of hearing difficulties, but these results, although positive as they highlight the suitability of ASH as a tool to describe hearing-related health states, may only be part of the picture. It is unknown whether these effects would be magnified further if individuals are provided with a reference ASH of “healthy” hearing, or if they are prompted to engage with the ASH using an active or passive engagement task. Does an active or passive engagement with ASH affect valuations for different health states, i.e. SiS and SiS+B? The hypothesis being that the addition of an active task will affect valuations obtained and lead to significant difference between listening situations and presentation methods, with active tasks leading to the most significant difference between SiS and SiS+B valuations collected using ASH alone.

The influence of a task on the long-term reliability of valuations will also be explored within this experiment. Does an individual rating a health state described using ASH need to rate the health state multiple times for the valuations to be stable and reliable? How many times should someone rate a health state described using ASH if more than one presentation of ASH are required for valuations to be reliable? Would multiple presentations and active engagement influence or affect the reliability of health states. The hypothesis being that engaging with the simulations will increase the reliability of valuations gained, in line with the findings of Goldstein et al, meaning fewer presentations will be needed to achieve optimum reliability.

5.3.1.1 Summary of research questions

Experiment 4 aims to address the following research questions:

1. Will there be an effect of engagement on health state valuation values

2. Will there be an effect of environment when participants are not provided with an example of normal hearing

5.3.2 Methods

5.3.2.1 The type of task used

When addressing the matter of which tasks to use within the experiment to promote engagement with the simulations, there were numerous options available. Initial thoughts for potential tasks included; listening for a target word, listening for the n^{th} word spoken by a target speaker, and questions with multiple choice answers.

Listening for a target word, or for the n^{th} word presented by a certain speaker, poses a risk that the individual rating the ASH could cease to listen to the simulations after they believe they have heard the n^{th} word, often referred to as attentional blink (Anderson, 2005). This could lead to the participants missing the simulations main facet, to communicate hearing difficulties associated with SSD, such as difficulty localising sound. There is also the issue that a participant could misidentify the target word or n^{th} word, they could hear a word with a similar phoneme or a similar speaker, and not realise that an individual with SSD could not localise the speaker in such a situation because the participant is not listening to the simulation as a whole. Using an n^{th} word may also lead to inattentional blindness, a phenomenon where participants would only listen for the target word and miss important information within the simulations (Neill et al., 1995, Simons, 2000).

Another type of task that could have been used to promote engagement is some form of multiple-choice question, such as “the target talker said…” and several options are then presented. This again poses the risk that the participant could only partially engage with the simulation as they could guess the correct answer, or again missing the message of the ASH, that localising one person within a group of speakers is difficult without normal hearing, especially if the answers presented were sentences spoken within the simulations.

A similar task to the multiple-choice question was also considered, a task where the participant had to select an audio example of the target speaker.

This task would have been inappropriate as again the participant could have guessed. Also, there are a limited number of potential speakers present; as there are not a large selection of speakers from within the IEEE and BKB sentence corpuses and the set-up of the experiments necessitate multiple presentations of ASH to be used within the experiment, this could have caused confusion and repetition of speakers, which may in turn have affected the levels of attention and engagement participants presented.

Previously in pilots 2 and 3, and in experiments 2 and 3, participants performed a listening task to validate the ASH's ability to recreate the hearing difficulties associated with SSD. During these experiments' participants were instructed to listen to the talker sat directly in front of them and note down the third sentence the talker said. This method has several advantages; it has been used within previous experiments, there are several demonstrable examples of its practical application with ASH, which is still a relatively unknown and foreign medium. The task is straightforward and can be explained efficiently to participants using written instructions. The task directs general attention to the target talker without influencing or biasing the participant by telling them to pay attention to specific details within the ASH that they may otherwise have missed. It would be advantageous to use this task again for the active engagement task as its use has been practically demonstrated and its suitability proven by previous experiments. Therefore, the listening task for experiments 4 and 5 will also utilise the task used within pilots 2 and 3 and experiments 2 and 3.

With the task selected there were still many questions surrounding the provision of feedback regarding the participants' performance at the listening task. Within this task the purpose of feedback was not for participants to use said feedback to improve their performance but so that they could use it as a way of quantifying how well they "heard" when listening to the ASH. It was important that the feedback given provided a level of detail that the participant could quickly use to evaluate the performance of someone with SSD, but not too in depth that it took focus away from the health valuation process. The feedback also needed to be realistic and pragmatic to administer.

The simplest and most straightforward way of providing this within the experiment would be for the researcher to provide feedback, in real time as the tasks were being completed by the participant. However, this would be impractical if the methods were scaled up and an economic evaluation were carried out with a large number of participants, as the time commitment required to prepare and then give the feedback within sessions would be immense, and as the researcher would have to be in the room with the participant this also introduces the potential for researcher or observer bias.

A less onerous method to provide feedback would be to provide it digitally. When the listening task was used in previous experiments, performance within the task was measured by the percentage of key words correctly identified. The feedback could be given by displaying the percentage of words correctly identified from the target sentence once the participant has entered the sentence, but this risks the participant’s attention switching to their performance percentage, or misinterpreting the main message of the ASH, the problems individuals with SSD have with localising sound. It is important that participants rating the health of people with SSD using the ASH have a complete picture of health state and the impact of SSD on quality of life.

Therefore, it could be better to passively provide feedback to participants, so to offer response that shows their performance at the listening task without directing their attention away from the valuation process. The easiest solution to providing feedback in this way is to let the participants make their own inferences about how the SSD is affecting their listening performance and how this would in turn affect their quality of life and well-being. This is similar in process to how a participant would value a health state by reading a vignette of a health state. See figure 15 for an example of the listening task and sentence reveal for the active listening condition.

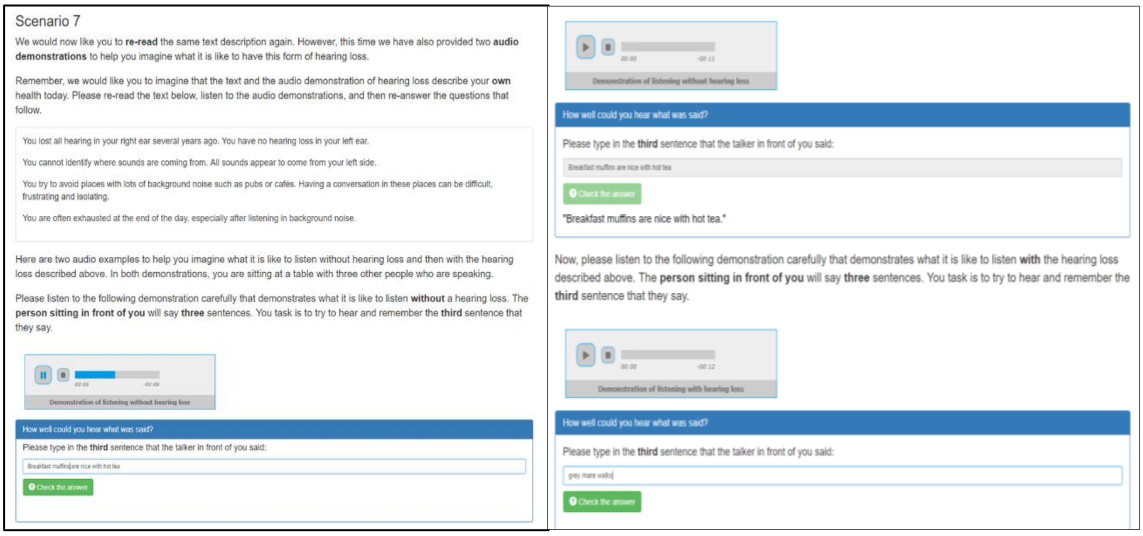


Figure 15: Active listening task; left: before sentence reveal, right: after sentence reveal

5.3.2.2 Participants

This experiment was conducted in accordance with the permissions granted by the NHS Grampian Research Ethics Committee and the Sponsor (University of Nottingham). Written informed consent was obtained for all participants by the Principal Investigator. 32 normal hearing participants (8 males, 24 females) with no expert knowledge of hearing research or audiology took part in the study. Participants had an age range of 21 to 44 years old (mean age 24.9 years) and all participants had pure-tone average threshold ≤ 30 dB HL [Hearing Level] in both ears and did not have a current ear infection or any other condition that would prohibit a hearing assessment taking place prior to the testing session. Participants were recruited via an email sent within the University of Nottingham Medical School, via the public and participant engagement website ‘Call for participants’ and via placed in student areas on the University of Nottingham University park campus. Participants were financially compensated for their time at a rate of £5 per hour. The experiment lasted approximately 45 minutes.

5.3.2.3 Counterbalancing

As with previous experiments (see 4.5.2.3) to control for order effects the conditions in Experiment 4 were counterbalanced using a Latin Square design (Graziano and Raulin, 2014). A 4x4 Latin Square was used to counterbalance the conditions presented to each participant. The 4x4 Latin Square sequence was repeated for every fourth participant to ensure that the specific list of sentences used for talker A were not repeated within each participant's testing session. Complete counterbalancing was used so that all possible orders of conditions occurred, that all conditions were presented an equal number of times in each position in the testing order, and each condition preceded and followed each other condition an equal number of times.

5.3.2.4 Sample size calculation

To ensure statistical power was reached for experiment 1 (see 3.4.2.1) data collected from pilot 1 was used to perform a priori power analysis, the sample size computed varied between 8 (based on the effect size of 1.04 for A alone SiS group and text alone SiS group data from pilot 1, with the required power level 0.80) and 16 (based on the effect size of 0.62 for A alone SiS+B group and text alone SiS+B group with the required power level 0.80). Experiment 4 was powered based on data from previous experiments, using the smallest effect size observed from experiment 1. A power analysis for a repeated measures ANOVA with two groups was conducted in G*Power (Faul 2007, 2009) to determine a sufficient sample size using an alpha of 0.05, a power of 0.80, and the effect size for the A alone SiS group within experiment 1 (f=0.47), the effect size for the A alone SiS+B group within experiment 1 (f=0.52), and a medium effect size (f = 0.7) (Faul et al., 2013). Based on these assumptions, the desired sample size was 8 based on the effect size for the A alone SiS group (0.42) and the A alone SiS+B group within experiment 1, and 6 based on a medium effect size (0.7). As these sample sizes were very manageable and lower than the sample sizes of previous experiments it was decided that a target recruitment figure of 16 to maintain comparability and simulate in recruitment figures across experiments.

5.3.2.5 Design

The study was a 2x2 factorial, between measures design. Four conditions were defined from the combination of: the two listening environments (SiS and SiS+B) and the two engagement types (active and passive). The simulations presented in Experiment 4 were version 2 simulations, the same simulations as those used in pilot 2 and 3 and experiment 3, as previously detailed in Chapter 3 sections 4.3, 4.4 & 4.5.

5.3.2.6 Procedure

As with previous experiments, participants gave informed consent to participate in the study. Once consent was obtained a hearing screening test was completed. The hearing screening test was carried out by the researcher on the day of the study and a screening audiogram was produced to ensure participants had a PTA threshold ≤ 30 dB HL [Hearing Level] in both ears.

Simulations were calibrated using a Bruel and Kjær Type 2250 sound pressure level meter which was attached to an artificial ear in which a pressure field ½ “ type 4192 microphone was used (see section 3.3.2.3, figure 7). The sound level meter was calibrated against a 94dB sound pressure level 1 kHz tone reference, type 4231. The Beyerdynamic DT 770 Pro Studio closed back headphones were placed over the artificial ear and ten simulations were played, the average sound level was taken across all simulations. Simulations were presented at 65.2 dB in the left ear, normal hearing ear, and 0 dB in the ear with simulated profound deafness. Simulations were presented within the 1 dB tolerance level set by the researcher. A level of 65 dB was chosen as this is the sound level for normal speech.

As with all previous experiments’ participants were seated at a computer situated within a soundproof booth and asked to put on a pair of closed back Beyerdynamic DT 770 Pro Studio headphones so that they comfortably enclosed their pinnae. Participants were presented with an example of the task to come, asked to rate reference health states (healthy, deaf and worst imaginable) and to rate five SiS and five SiS+B health states presented using ASH and valued using TTO and VAS tasks.

Participants within the passive listening condition were asked to listen to ASH presented binaurally (the simulations which simulated normal hearing) and then monaurally (the simulations which simulated an SSD health state) in their entirety, before rating the health state described. Participants within the active listening condition were asked to listen to ASH presented binaurally (the simulations which simulated normal hearing) and to complete the behavioural task in which they were asked to write down the third sentence spoken by the talker at position A. Participants were then asked to listen to ASH presented monaurally (the simulations which simulated a SSD health state) in their entirety and to complete the behavioural task in which they were asked to write down the third sentence spoken by the talker at position A, they were then asked to rate the health state that had been described. Participants would then carry out the TTO and VAS valuation tasks for each SiS and SiS+B presentation. Binaural and monaural simulations presented within the same scenario (for example the Scenario 7 presentations shown in figure 15) were presented using different ASH with different target talkers and sentences. This ensured that participants could not use the binaural simulations as a reference when completing the behavioural task presented for the monoaural ASH.

5.3.2.7 Analysis

The TTO data set was automatically transformed into a utility value by the web interface, using the formula also detailed in Chapter 2 (see section 2.5.2). The VAS data were transformed into utility values (U VAS) using the formula previously detailed in Chapter 2 (see section 2.5.3). Results from Experiment 4 relating to research question one were analysed using ANOVA and IIC values. Results relating to research question two looked at the reliability of the utility values obtained and so were analysed using ICC analysis.

5.3.3 Results

The health utility results collected using TTO method are shown in Table 12 for the active engagement condition and are shown in Table 13 for the passive engagement condition. The health utility results collected using VAS method are shown in Table 14 for the active engagement condition and are shown in Table 15 for the passive engagement condition.

As with previous studies within this work the average utility value for the active engagement TTO perfect health was close to 1 (healthy = 0.96) (see section 2.2), the average score for perfect health with deafness was in line with figures previously found in the literature (see section 2.4.1) (deaf = 0.70) and average score for worst health imaginable was close to 0 (see section 2.2) (worst = 0.14). The average utility value for the passive engagement TTO perfect health was close to 1 (healthy = 0.97), the average score for perfect health with deafness was close to figures previously found in the literature (deaf = 0.62) and average score for worst health imaginable was close to 0 (worst = 0.10).

The average utility value for the active engagement VAS perfect health was close to 1 (healthy = 0.97), the average score for perfect health with deafness was higher than the value reported in previous literature (deaf = 0.91) and average score for worst health imaginable was close to 0 (worst = 0.35). The average utility value for the passive engagement VAS perfect health was close to 1 (healthy = 0.99), again the average score for perfect health with deafness was higher than figures previously found in the literature (deaf = 0.84) and average score for worst health imaginable was close to 0 (worst = 0.20).

The average SiS and average SiS+B TTO valuations for each participant were subjected to a repeated measures ANOVA with factors of environment (SiS vs SiS+B), and engagement (active vs passive). The main effect of environment ($F(1,30)= 49.70$, $p<.00$) was significant. There was not a significant interaction between environment and engagement ($F(1,30)= 0.026$, $p=0.874$) environment and repetition ($F(1,15)= .775$, $p=.546$), and engagement and repetition ($F(1,15)= .258$, $p=.904$). The average SiS and average SiS+B VAS valuations for each participant were also subjected to a repeated measures ANOVA with factors of environment (SiS vs SiS+B), and engagement (active vs passive). The main effect of environment ($F(1,30)= 10.03$, $p<.05$) was significant. There was not a significant interaction between environment and engagement ($F(1,30)= 2.12$, $p=0.156$).

ICC values were calculated using the repeated valuations of the same health state description within the experiment (SiS valuations and SiS+B valuations) for the TTO values (see Table 12 and 13 and figure 16). The ICC values for all conditions were above 0.75 and are considered stable, however only the ICC value for the active engagement SiS+B condition had a confidence interval that did not pass below the 0.75 threshold for acceptability (see section 3.5.2.2). ICC values were also calculated for the VAS values (see Table 14 and 15 and figure 17). Only the ICC value for Passive SiS+B condition was above 0.75 and had a confidence interval that did not pass below the 0.75 threshold for acceptability and so considered stable (see section 3.5.2.2).

Participant	Healthy	Deaf	Worst	SiS 1	SiS 2	SiS 3	SiS 4	SiS 5	SiS+B 1	SiS+B 2	SiS+B 3	SiS+B 4	SiS+B 5
1	0.90	0.80	0.30	0.75	0.80	0.76	0.79	0.75	0.75	0.70	0.72	0.68	0.73
2	1.00	0.59	0.01	0.00	0.31	0.30	0.40	0.50	0.30	0.19	0.30	0.21	0.31
3	0.94	0.99	0.64	0.88	0.89	0.86	0.80	0.70	0.92	0.78	0.71	0.69	0.68
4	0.90	0.60	0.01	0.69	0.42	0.52	0.54	0.49	0.50	0.39	0.32	0.39	0.33
5	1.00	0.40	0.00	0.30	0.21	0.14	0.13	0.19	0.20	0.17	0.15	0.09	0.14
6	1.00	0.80	0.39	0.60	0.55	0.50	0.49	0.50	0.50	0.60	0.40	0.41	0.21
7	0.99	0.70	0.20	0.80	0.69	0.90	0.90	0.84	0.60	0.60	0.80	0.41	0.60
8	0.99	0.49	0.01	0.90	0.95	0.90	0.90	0.83	0.86	0.80	0.94	0.83	0.79
9	0.95	0.81	0.05	0.60	0.40	0.41	0.38	0.50	0.38	0.31	0.48	0.44	0.31
10	1.00	0.95	0.30	0.80	0.90	0.67	0.50	0.54	0.79	0.69	0.60	0.50	0.54
11	0.99	0.60	0.04	0.70	0.70	0.50	0.36	0.21	0.60	0.40	0.60	0.40	0.40
12	0.90	0.60	0.01	0.70	0.80	0.80	0.60	0.79	0.60	0.70	0.59	0.70	0.60
13	1.00	0.79	0.01	0.30	0.20	0.20	0.20	0.10	0.20	0.20	0.20	0.10	0.11
14	1.00	0.70	0.00	0.40	0.50	0.30	0.40	0.60	0.50	0.50	0.30	0.20	0.20
15	0.90	0.60	0.20	0.40	0.50	0.30	0.40	0.60	0.40	0.35	0.40	0.30	0.20
16	0.97	0.70	0.02	0.20	0.30	0.30	0.36	0.30	0.30	0.30	0.20	0.25	0.25
Average	0.96	0.70	0.14	0.56	0.57	0.52	0.51	0.53	0.53	0.48	0.48	0.41	0.40

Table 12: Transformed TTO data for Experiment 4 Active condition

Participant	Healthy	Deaf	Worst	SiS 1	SiS 2	SiS 3	SiS 4	SiS 5	SiS+B 1	SiS+B 2	SiS+B 3	SiS+B 4	SiS+B 5
1	0.94	0.65	0.01	0.84	0.90	0.82	0.79	0.92	0.84	0.76	0.84	0.62	0.89
2	1.00	0.60	0.20	0.80	0.80	0.80	0.80	0.80	0.80	0.70	0.60	0.60	0.60
3	1.00	0.50	0.02	0.80	0.31	0.30	0.30	0.50	0.20	0.30	0.30	0.29	0.30
4	0.96	0.60	0.02	0.97	0.94	0.91	0.88	0.94	0.65	0.70	0.90	0.95	0.90
5	0.99	0.99	0.11	0.59	0.50	0.60	0.56	0.47	0.50	0.49	0.55	0.60	0.41
6	0.99	0.92	0.69	0.75	0.73	0.71	0.70	0.73	0.69	0.63	0.59	0.62	0.53
7	0.95	0.57	0.00	0.80	0.68	0.69	0.72	0.79	0.74	0.62	0.58	0.54	0.61
8	0.91	0.62	0.07	0.60	0.53	0.56	0.50	0.48	0.64	0.71	0.63	0.54	0.49
9	0.90	0.20	0.00	0.80	0.80	0.70	0.70	0.70	0.60	0.60	0.50	0.68	0.50
10	0.90	0.69	0.26	0.68	0.70	0.67	0.75	0.72	0.64	0.70	0.78	0.67	0.72
11	1.00	0.60	0.00	0.80	0.64	0.60	0.70	0.66	0.65	0.60	0.66	0.70	0.70
12	0.97	0.65	0.02	0.80	0.83	0.83	0.83	0.76	0.80	0.74	0.61	0.73	0.77
13	1.00	0.55	0.00	0.40	0.35	0.30	0.30	0.30	0.30	0.30	0.24	0.23	0.20
14	1.00	0.70	0.00	0.40	0.50	0.46	0.46	0.50	0.55	0.40	0.46	0.44	0.40
15	1.00	0.60	0.11	0.54	0.56	0.59	0.56	0.56	0.63	0.56	0.52	0.58	0.50
16	1.00	0.50	0.01	0.77	0.77	0.76	0.74	0.89	0.75	0.65	0.56	0.56	0.54
Average	0.97	0.62	0.10	0.71	0.66	0.64	0.64	0.67	0.62	0.59	0.58	0.58	0.57

Table 13: Transformed TTO data for Experiment 4 Passive condition.

Participant	Healthy	Deaf	Worst	SiS 1	SiS 2	SiS 3	SiS 4	SiS 5	SiS+B 1	SiS+B 2	SiS+B 3	SiS+B 4	SiS+B 5
1	1.00	0.99	0.67	0.97	1.00	1.00	1.00	1.00	0.97	0.99	0.99	0.97	0.99
2	1.00	0.85	0.56	1.00	0.43	0.77	0.85	0.68	0.67	0.43	0.56	0.42	0.42
3	1.00	1.00	0.16	0.97	0.92	0.30	0.16	0.50	0.97	0.42	0.18	0.30	0.38
4	0.99	0.92	0.43	0.97	0.92	0.98	0.97	0.97	0.85	0.67	0.67	0.85	0.77
5	1.00	0.85	0.30	0.67	0.54	0.36	0.51	0.59	0.56	0.43	0.43	0.43	0.49
6	0.85	0.92	0.67	0.77	0.67	0.77	0.77	0.85	0.43	0.43	0.85	0.55	0.76
7	1.00	0.77	0.43	0.85	0.67	1.00	0.99	0.97	0.68	0.56	0.78	0.42	0.76
8	1.00	0.97	0.67	1.00	1.00	1.00	1.00	1.00	0.99	0.97	1.00	0.99	0.99
9	1.00	0.99	0.16	0.67	0.42	0.67	0.67	0.77	0.56	0.56	0.77	0.77	0.67
10	1.00	1.00	0.92	0.92	0.97	0.89	0.85	0.85	0.97	0.89	0.89	0.86	0.89
11	1.00	0.56	0.04	0.67	0.81	0.43	0.43	0.30	0.43	0.37	0.77	0.30	0.37
12	1.00	0.97	0.00	1.00	1.00	1.00	0.97	1.00	0.97	1.00	0.97	1.00	0.97
13	1.00	0.92	0.04	0.85	0.92	0.92	0.92	0.67	0.77	0.92	0.92	0.77	0.67
14	1.00	0.92	0.16	0.56	0.67	0.56	0.56	0.77	0.67	0.67	0.43	0.37	0.30
15	0.67*	0.92	0.30	0.56	0.67	0.56	0.56	0.77	0.85	0.81	0.77	0.85	0.67
16	1.00	0.92	0.16	0.77	0.77	0.77	0.67	0.77	0.67	0.77	0.85	0.77	0.67
Average	0.97	0.91	0.35	0.83	0.78	0.75	0.74	0.78	0.75	0.68	0.74	0.67	0.67

Table 14: Transformed VAS data for Experiment 4 Active condition

Participant	Healthy	Worst	Deaf	SiS 1	SiS 2	SiS 3	SiS 4	SiS 5	SiS+B 1	SiS+B 2	SiS+B 3	SiS+B 4	SiS+B 5
1	1.00	0.67	0.00	0.93	0.85	0.85	0.85	0.92	0.92	0.92	0.93	0.85	0.85
2	1.00	0.77	0.43	0.72	0.77	0.79	0.75	0.56	0.68	0.64	0.61	0.62	0.62
3	1.00	0.92	0.56	1.00	1.00	1.00	1.00	0.99	1.00	1.00	0.99	0.99	0.99
4	1.00	0.99	0.16	0.99	0.99	0.99	0.99	0.99	0.97	0.97	0.97	0.97	0.97
5	0.99	0.30	0.00	0.97	0.92	0.92	0.92	0.97	0.85	0.92	0.92	0.92	0.85
6	0.92	0.92	0.78	0.67	0.67	0.54	0.57	0.56	0.84	0.55	0.63	0.45	0.56
7	1.00	0.43	0.00	0.58	0.67	0.55	0.54	0.63	0.66	0.92	0.46	0.58	0.85
8	1.00	0.92	0.56	0.66	0.78	0.67	0.85	0.77	0.85	0.66	0.43	0.75	0.76
9	0.00	0.00	0.43	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.43	0.00
10	1.00	1.00	0.16	0.97	0.92	0.92	0.92	0.92	1.00	0.93	0.92	0.92	0.92
11	1.00	0.92	0.67	0.97	0.95	0.93	0.94	0.96	0.93	0.90	0.90	0.89	0.89
12	1.00	0.77	0.16	0.92	0.97	0.99	0.97	0.97	0.95	0.89	0.85	0.95	0.89
13	1.00	1.00	0.43	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
14	1.00	0.77	0.16	0.97	0.97	0.96	0.99	0.99	0.97	0.95	0.95	0.93	0.93
15	1.00	0.97	0.00	1.00	1.00	1.00	0.97	0.97	1.00	1.00	1.00	1.00	0.97
16	1.00	0.88	0.56	1.00	1.00	0.97	0.97	0.99	1.00	0.92	0.92	0.92	0.85
Average	0.93	0.77	0.32	0.84	0.84	0.82	0.83	0.83	0.85	0.82	0.78	0.82	0.81

*Table 15: Pearson's r values for TTO and VAS SiS and SiS+B values. * Correlation significant to 0.05 level. ** Correlation significant to the 0.01 level.*

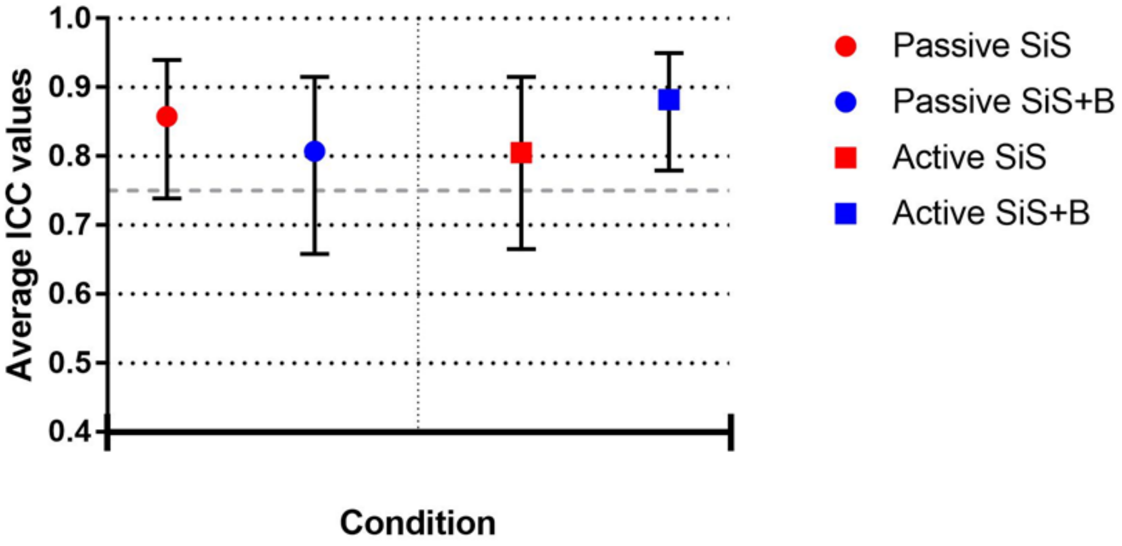


Figure 16: Mean intra-class correlation coefficients (ICCs) and their 95% confidence intervals for TTO data from Experiment 4.

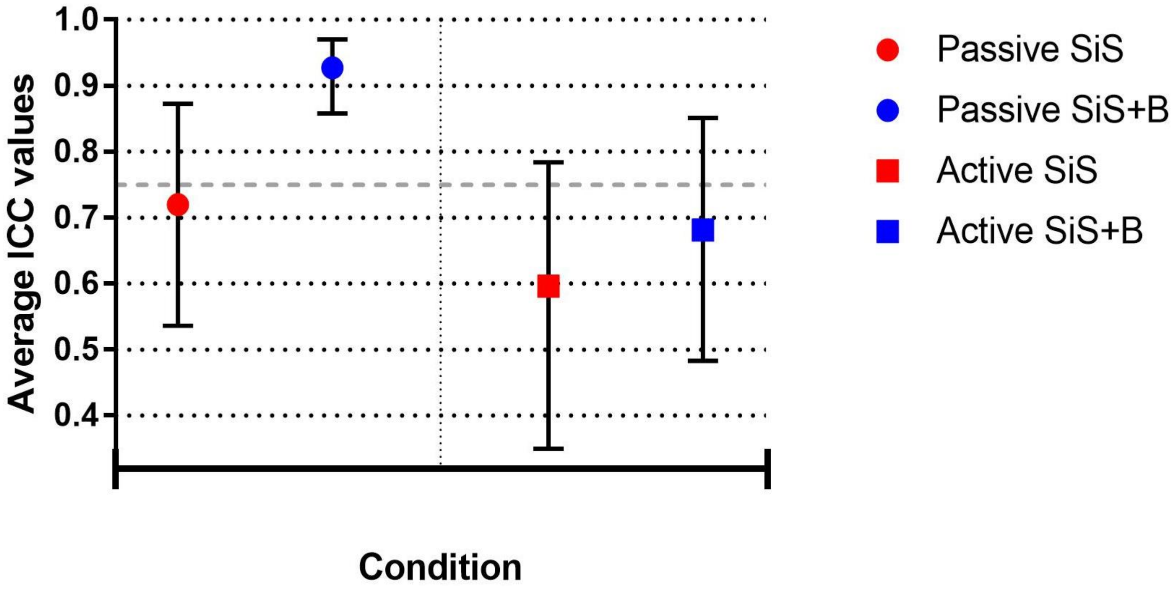


Figure 17 : Mean intra-class correlation coefficients (ICCs) and their 95% confidence intervals for VAS data from Experiment 4

5.3.4 Discussion

The reference health states collected indicate that the participants' valuations were in line with the valuations collected from previous studies and that the main utility task, the TTO task, and the secondary utility collection task, the VAS task, were generally understood. Again, within experiment 4 there was a significant effect of environment, meaning that participants viewed the SiS and SiS+B health states as representing different levels of difficulty, regardless of whether active or passive engagement was facilitated. Overall there was no effect of engagement on the utility values collected, this could indicate that adding an active task to the valuation process when using ASH may not influence participants' valuations. Before surmising the usefulness and influence of an active engagement task in the health utility valuation process it would be pertinent to clarify whether this is still the case for ASH that are presented without a comparator of 'healthy' hearing.

5.4 Experiment 5: Is an active or passive task needed to promote engagement with ASH if no ASH of 'healthy hearing' is presented and a disease only ASH is presented?

5.4.1 Introduction

Experiment 5 investigated whether there was still a difference in the valuations given for each listening environment (SiS and SiS+B), as found in previous experiments, when participants are also provided with an ASH demonstrating normal hearing, and to investigate the effect an active listening task has on health state valuations when participants are provided with an ASH demonstrating normal hearing. During initial piloting of version 1 simulations, specifically within pilot 1, a simulation demonstrating 'healthy' hearing was provided to participants. These 'healthy hearing' simulations were replications of the ASH of the monaural SSD hearing-related health state and used the same environment and speaker set up as the monaural ASH. As the ASH were new and untested, this healthy hearing simulation was given so that the participants had reference point for healthy and unhealthy in case the monaural ASH did not adequately convey the hearing difficulties associated with SSD and the different effects SSD has on an individual's ability to hear within different environments (SiS and SiS+B). After pilot 1 the provision of a healthy hearing ASH was not repeated, as the improvements to ASH between version 1 and version 2 deemed it unnecessary as the ASH were truer to life and much better at communicating listening difficulties. However, it is unclear if this decision was made too hastily, and without empirical evidence it cannot be categorically said that ASH need only be presented monaurally. Do participants still need a simulation of healthy hearing as a reference, even with the improvements to the ASH? Could the provision of a second simulation as a comparator be beneficial to participants' understanding of the impact of SSD on quality of life? Would such a comparator simulation make ASH easier to understand and consequently cause participants to change their mind less and produce more reliable results?

It could be the case that by not providing two simulations to highlight the difference between healthy hearing and a hearing impairment the participants miss or misinterpret the severity of the difficulties that the simulations are portraying.

Again, the question surrounding the effect of an active listening task are still present when both a monaural and binaural simulation are provided: does an active or passive engagement with ASH effect valuations for different health states, i.e. SiS and SiS+B when participants are also provided with an ASH demonstrating normal hearing? The hypothesis being that the addition of an active task, provided alongside a different ASH demonstrating normal hearing, will affect valuations obtained, and lead to significant difference between listening situations and presentation methods, with active tasks leading to the most significant difference between SiS and SiS+B valuations collected using ASH alone.

Also, within this experiment the influence of a listening task on the reliability of valuations collected when participants are also provided with an ASH of binaural hearing and monoaural hearing will be explored. Does an individual rating a health state described using binaural and monaural ASH and an active listening task need to rate the health state multiple times for the valuations to be stable and reliable? Experiment 4 indicated that valuations provided using ASH were stable regardless of the presence of an active and passive task, will this again be the case when a monaural and binaural simulation are provided. The hypothesis being that engaging with the simulations and attending to their content will increase the reliability of valuations gained, in a similar way that happened in the studies examining the effects of selective attention and attentional lapses (Schooler et al., 2004, Reason and Lucas, 1984, Manly et al., 1999) so less presentations will be needed to achieve optimum reliability.

5.4.1.1 Summary of research questions

Experiment 5 aimed to address the following research questions:

1. There will be an effect of engagement on health state valuation values when participants are provided with an example of normal hearing
2. There will be an effect of environment when participants are provided with an example of normal hearing

5.4.2 Methods

5.4.2.1 The type of task used

Experiment 5 used the same listening task as experiment 4, see section 5.2.2.1.

5.4.2.2 Participants

This experiment was conducted in accordance with the permissions granted by the NHS Grampian Research Ethics Committee and the Sponsor (University of Nottingham) Written informed consent was obtained for all participants by the Principal Investigator. Thirty-two normal hearing participants (9 males, 23 females) took part in the study. Participants had an age range of 22 to 36 years old (mean age 25.2 years) and all participants had pure-tone average threshold ≤ 30 dB HL [Hearing Level] in both ears and did not have a current ear infection or any other condition that would prohibit a hearing assessment taking place prior to the testing session. Participants were recruited via an email sent within the University of Nottingham Medical School, via the public and participant engagement website ‘Call for participants’ and via placed in student areas on the University of Nottingham University park campus. Participants were financially compensated for their time at a rate of £5 per hour. The experiment lasted approximately 45 minutes.

5.4.2.3 Counterbalancing

To control for order effects the conditions in Experiment 5 were counterbalanced as conditions were counterbalanced for experiment 4 (see 5.3.2.3).

5.4.2.4 Sample size calculations

Experiment 5 was powered using the power calculation performed experiment 4, and so had a target recruitment figure of 16 for each engagement condition. As with experiment 4 this figure was selected to account for possible attrition, and to maintain comparability and similitude to experiment 4 so that the two experiments can be easily compared and to maintain comparability across all other experiments within this body of work.

5.4.2.5 Design

The study was a 2x2 factorial, between measures design. Four conditions were defined from the combination of the two listening environments (SiS and SiS+B) and the two engagement types (active and passive). Each participant took part in one session. In the session, 5 simulations were presented in each condition, resulting in a total of 10 simulations being presented. The simulations presented in Experiment 5 were version 2 simulations, the same simulations as those used in pilot 2 and 3 and experiment 3 and 4, as previously detailed in Chapter 4 sections 4.3, 4.4 and 4.5. As in experiment 4 both the binaural and monaural simulations presented within a scenario were presented using different ASH with different target talkers and sentences, to ensure that participants could not use the binaural simulations as a reference when completing the behavioural task presented for the monoaural ASH.

5.4.2.6 Procedure

As with previous experiments participants involved in these experiments gave informed consent to participate in the study and completed a hearing screening test, which was carried out by the researcher on the day of the study. Once the consent was obtained a screening audiogram was obtained to ensure participants had a PTA threshold ≤ 30 dB HL [Hearing Level]. Simulations were calibrated using a Bruel and Kjær Type 2250 sound pressure level meter which was attached to an artificial ear in which a pressure field ½ “ type 4192 microphone was used (see figure 7 section 3.3.23). Simulations were presented at 65.2 dB in the left ear, normal hearing ear, and 0 dB in the ear with simulated profound deafness. Simulations were presented within the 1 dB tolerance level set by the researcher. A level of 65 dB was chosen as this is the sound level for normal speech.

As with all previous experiments' participants were seated at a computer situated within a soundproof booth and asked to put on a pair of closed back Beyerdynamic DT 770 Pro Studio headphones so that they comfortably enclosed their pinnae. Participants were presented with an example of the task to come, asked to rate the reference health states: healthy, deaf and worst health imaginable, they were then asked to rate five SiS and five SiS+B health states presented using ASH and a text excerpt, valued using TTO and VAS tasks. Participants within the passive listening condition were asked to listen to the ASH presented both binaurally and monaurally in their entirety, before rating the health state described. Participants within the active listening condition were asked to listen to the ASH presented both binaurally and monaurally in their entirety, they were then asked to complete the behavioural task, in which they were asked to type the third sentence spoken by the talker at position A into a blank text box and click a button on the screen to reveal the sentence spoken by the target talker. Participants were then asked to rate the health state that had been described using TTO and VAS valuation tasks. Both TTO AND VAS valuation tasks were performed for each SiS and SiS+B ASH presentation.

5.4.2.7 Analysis

The TTO data set was automatically transformed into a utility value using the formula also detailed in Chapter 2 (see section 2.5.2). The VAS data were transformed into utility values (U VAS) using the formula previously detailed in Chapter 2 (see section 2.5.3). Results from Experiment 5 relating to research question one were analysed using ANOVA and IIC values. Results relating to research question two looked at the reliability of the utility values obtained and so were analysed using ICC analysis.

5.4.3 Results

The health utility results collected using the TTO method are shown in Table 18 for the active engagement condition and are shown in Table 17 for the passive engagement condition. The health utility results collected using the VAS method are shown in Table 20 for the active engagement condition and are shown in Table 18 for the passive engagement condition.

The reference health states collected indicate that the participants were in line with the population in their judgements and that the main utility task, the TTO task, and the secondary utility collection task, the VAS task, were generally understood.

The average utility value for the active engagement TTO perfect health was close to 1 (healthy = 0.95) in line with previous studies within this work (see section 2.2), the average score for perfect health with deafness was in line with figures previously found in the literature (see section 2.4.1) (deaf = 0.60) and average score for worst health imaginable was close to 0 (see section 2.2) (worst = 0.06). The average utility value for the passive engagement TTO perfect health was close to 1 (healthy = 0. 97), the average score for perfect health with deafness was close to figures previously found in the literature (deaf = 0. 59) and average score for worst health imaginable was close to 0 (worst = 0. 10). The average utility value for the active engagement VAS perfect health reached the upper limit of measurement (healthy = 1.00), the average score for perfect health with deafness was higher than figures previously found in the literature (deaf = 0.86) and average score for worst health imaginable was close to 0 (worst = 0. 31). The average utility value for the passive engagement VAS perfect health was close to 1 (healthy = 0. 99) and average score for worst health imaginable was close to 0 (worst = 0. 20), again the average score for perfect health with deafness was higher than figures previously found in the literature (deaf = 0. 84).

The average SiS and average SiS+B TTO valuations for each participant were subjected to a repeated measures ANOVA with factors of environment (SiS vs SiS+B), and a between subject factor of engagement (active vs passive). The main effect of environment ($F(1,30)= 11.46$, $p=.002$) was significant and there was not a significant interaction between environment and engagement ($F(1,30)=.29$, $p<.60$). The average SiS and average SiS+B VAS valuations for each participant were subjected to a repeated measures ANOVA with a within factor of environment (SiS vs SiS+B), and a between group factor of engagement (active vs passive). The main effect of environment ($F(1,30)=6.07$, p=.02) was significant and there was not a significant interaction between environment and engagement (F(1,30)=2.21, p=.15).

ICC values were calculated for repeated valuations of the same health state description within the experiment (SiS valuations and SiS+B valuations) (see Table 21 and figure 18). The average ICC values for Passive SiS, Passive SiS+B and Active SiS+B conditions were above 0.75 and are considered stable, however, only the ICC value for the Active SiS+B condition had a confidence interval that did not pass below the 0.75 threshold. for acceptability (see section 3.5.2.2). ICC values were also calculated for the VAS values (see Table 21 and figure 19). All ICC values for all conditions collected using VAS were above the 0.75 threshold for acceptability and so considered stable, however only the ICC value for the Passive SiS, passive SiS+ and Active SiS conditions had a confidence interval that did not pass below the 0.75 threshold for acceptability (see section 3.5.2.2).

Participant	Healthy	Worst	Deaf	SiS 1	SiS 2	SiS 3	SiS 4	SiS 5	SiS+B 1	SiS+B 2	SiS+B 3	SiS+B 4	SiS+B 5
1	0.99	0.60	0.00	0.69	0.60	0.54	0.70	0.65	0.56	0.80	0.70	0.60	0.55
2	1.00	0.31	0.01	0.25	0.27	0.30	0.33	0.28	0.33	0.26	0.23	0.24	0.29
3	0.99	0.60	0.10	0.96	0.92	0.92	0.89	0.74	0.96	0.72	0.75	0.73	0.73
4	1.00	0.70	0.00	0.75	0.75	0.80	0.84	0.85	0.65	0.65	0.65	0.65	0.65
5	0.90	0.20	0.01	0.79	0.80	0.70	0.69	0.79	0.60	0.60	0.59	0.70	0.50
6	0.81	0.40	0.49	0.98	0.59	0.80	0.69	0.99	0.99	0.98	0.98	0.91	0.99
7	1.00	0.39	0.01	0.40	0.46	0.50	0.45	0.41	0.31	0.16	0.62	0.51	0.33
8	1.00	0.50	0.03	0.52	0.31	0.30	0.50	0.41	0.70	0.41	0.32	0.24	0.27
9	0.99	0.81	0.11	0.90	0.89	1.00	0.80	0.90	1.00	0.90	0.80	0.70	0.90
10	0.99	0.70	0.07	0.62	0.32	0.29	0.28	0.25	0.80	0.31	0.24	0.24	0.25
11	0.96	0.50	0.16	0.75	0.50	0.55	0.61	0.60	0.40	0.45	0.55	0.52	0.51
12	0.95	0.70	0.10	0.74	0.85	0.84	0.86	0.89	0.83	0.67	0.56	0.73	0.75
13	1.00	0.90	0.00	0.95	0.96	0.91	0.85	0.94	0.88	0.75	0.86	0.88	0.82
14	0.98	0.60	0.01	0.82	0.82	0.76	0.80	0.87	0.80	0.75	0.80	0.70	0.71
15	1.00	0.90	0.30	1.00	0.90	1.00	0.70	0.75	0.90	0.90	0.85	0.85	0.65
16	0.99	0.60	0.16	0.90	0.90	0.79	0.79	0.83	0.97	0.79	0.70	0.69	0.59
Average	**0.97**	**0.59**	**0.10**	**0.75**	**0.68**	**0.69**	**0.67**	**0.70**	**0.73**	**0.63**	**0.64**	**0.62**	**0.59**

Table 16: Transformed TTO data for Experiment 5 passive condition

Participant	Healthy	Worst	Deaf	SiS 1	SiS 2	SiS 3	SiS 4	SiS 5	SiS+B 1	SiS+B 2	SiS+B 3	SiS+B 4	SiS+B 5
1	0.80	0.69	0.20	0.20	0.25	0.24	0.25	0.19	0.21	0.15	0.17	0.00	0.13
2	0.59	0.50	0.10	0.50	0.50	0.46	0.45	0.50	0.40	0.40	0.37	0.30	0.26
3	1.00	0.31	0.02	0.85	0.44	0.65	0.43	0.23	0.39	0.24	0.22	0.26	0.21
4	1.00	0.80	0.00	0.95	0.95	0.98	0.96	0.98	0.90	0.90	0.94	0.90	0.94
5	0.93	0.75	0.05	0.80	0.84	0.84	0.95	0.91	0.80	0.88	0.84	0.85	0.83
6	0.99	0.60	0.01	0.50	0.50	0.40	0.40	0.50	0.40	0.30	0.30	0.40	0.36
7	1.00	0.85	0.00	0.90	0.00	0.84	0.97	0.00	0.90	0.80	0.84	0.85	0.94
8	1.00	0.40	0.00	0.59	0.50	0.50	0.59	0.53	0.30	0.44	0.36	0.31	0.10
9	1.00	0.82	0.00	0.80	0.82	0.76	0.87	0.77	0.78	0.76	0.79	0.77	0.77
10	1.00	0.50	0.10	0.79	0.80	0.79	0.80	0.79	0.80	0.80	0.70	0.00	0.69
11	0.94	0.50	0.19	0.40	0.50	0.51	0.20	0.50	0.50	0.33	0.18	0.06	0.20
12	0.95	0.74	0.24	0.90	0.89	0.83	0.75	0.84	0.87	0.66	0.65	0.64	0.67
13	0.96	0.50	0.04	0.50	0.39	0.40	0.29	0.40	0.30	0.21	0.31	0.38	0.40
14	0.99	0.60	0.01	0.70	0.95	0.95	0.70	0.95	0.80	0.84	0.90	0.90	0.75
15	1.00	0.70	0.00	0.19	0.20	0.30	0.10	0.10	0.20	0.19	0.25	0.10	0.10
16	1.00	0.40	0.00	0.15	0.11	0.08	0.10	0.10	0.15	0.10	0.10	0.10	0.10
Average	**0.95**	**0.60**	**0.06**	**0.61**	**0.54**	**0.60**	**0.55**	**0.52**	**0.54**	**0.50**	**0.50**	**0.43**	**0.47**

Table 17 : Transformed TTO data for Experiment 5 active condition

Participant	Healthy	Worst	Deaf	SiS 1	SiS 2	SiS 3	SiS 4	SiS 5	SiS+B 1	SiS+B 2	SiS+B 3	SiS+B 4	SiS+B 5
1	1.00	0.67	0.00	0.93	0.85	0.85	0.85	0.92	0.92	0.92	0.93	0.85	0.85
2	1.00	0.77	0.43	0.72	0.77	0.79	0.75	0.56	0.68	0.64	0.61	0.62	0.62
3	1.00	0.92	0.56	1.00	1.00	1.00	1.00	0.99	1.00	1.00	0.99	0.99	0.99
4	1.00	0.99	0.16	0.99	0.99	0.99	0.99	0.99	0.97	0.97	0.97	0.97	0.97
5	0.99	0.30	0.00	0.97	0.92	0.92	0.92	0.97	0.85	0.92	0.92	0.92	0.85
6	0.92	0.92	0.78	0.67	0.67	0.54	0.57	0.56	0.84	0.55	0.63	0.45	0.56
7	1.00	0.43	0.00	0.58	0.67	0.55	0.54	0.63	0.66	0.92	0.46	0.58	0.85
8	1.00	0.92	0.56	0.66	0.78	0.67	0.85	0.77	0.85	0.66	0.43	0.75	0.76
9	0.00	0.00	0.43	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.43	0.00
10	1.00	1.00	0.16	0.97	0.92	0.92	0.92	0.92	1.00	0.93	0.92	0.92	0.92
11	1.00	0.92	0.67	0.97	0.95	0.93	0.94	0.96	0.93	0.90	0.90	0.89	0.89
12	1.00	0.77	0.16	0.92	0.97	0.99	0.97	0.97	0.95	0.89	0.85	0.95	0.89
13	1.00	1.00	0.43	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
14	1.00	0.77	0.16	0.97	0.97	0.96	0.99	0.99	0.97	0.95	0.95	0.93	0.93
15	1.00	0.97	0.00	1.00	1.00	1.00	0.97	0.97	1.00	1.00	1.00	1.00	0.97
16	1.00	0.88	0.56	1.00	1.00	0.97	0.97	0.99	1.00	0.92	0.92	0.92	0.85
Average	0.93	0.77	0.32	0.84	0.84	0.82	0.83	0.83	0.85	0.82	0.78	0.82	0.81

Table 18: Transformed VAS data for Experiment 5 passive condition

Participant	Healthy	Worst	Deaf	SiS 1	SiS 2	SiS 3	SiS 4	SiS 5	SiS+B 1	SiS+B 2	SiS+B 3	SiS+B 4	SiS+B 5
1	1.00	1.00	0.30	0.16	0.29	0.16	0.23	0.16	0.16	0.04	0.09	1.00	0.07
2	0.97	0.97	0.85	0.99	0.97	0.97	0.95	0.99	0.99	0.93	0.93	0.92	0.92
3	1.00	0.89	0.16	0.97	0.95	0.97	0.92	0.85	0.92	0.85	0.89	0.89	0.85
4	1.00	0.92	0.08	1.00	0.99	1.00	1.00	1.00	0.99	0.99	0.99	0.99	1.00
5	1.00	1.00	0.05	0.99	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00
6	1.00	0.97	0.16	0.85	0.92	0.85	0.85	0.97	0.67	0.77	0.67	0.85	0.85
7	1.00	0.99	0.97	0.99	1.00	0.97	1.00	1.00	0.99	0.97	0.99	0.99	0.99
8	1.00	0.67	0.00	0.92	0.67	0.67	0.77	0.85	0.56	0.85	0.66	0.43	0.16
9	1.00	1.00	0.30	0.97	0.99	0.99	1.00	0.99	0.97	0.99	0.99	0.99	0.99
10	1.00	0.67	0.16	0.97	0.92	0.97	0.92	0.92	0.97	0.92	0.92	1.00	0.85
11	1.00	0.67	0.92	0.83	0.66	0.66	0.29	0.67	0.67	0.48	0.24	0.17	0.30
12	1.00	0.92	0.30	0.98	0.98	0.93	0.90	0.93	0.95	0.84	0.77	0.77	0.80
13	1.00	0.93	0.45	0.85	0.67	0.84	0.61	0.66	0.70	0.59	0.85	0.76	0.70
14	1.00	0.92	0.16	0.85	1.00	1.00	0.85	1.00	0.92	0.95	0.99	1.00	0.92
15	1.00	0.92	0.16	0.37	0.38	0.56	0.30	0.16	0.38	0.37	0.43	0.30	0.16
16	1.00	0.30	0.01	0.30	0.30	0.16	0.16	0.16	0.30	0.16	0.16	0.16	0.16
Average	1.00	0.86	0.31	0.81	0.79	0.80	0.73	0.77	0.76	0.73	0.72	0.76	0.67

Table 19: Transformed VAS data for Experiment 5 active condition

		Condition			
		Passive SiS	**Passive SiS+B**	**Active SiS**	**Active SiS+B**
Valuation method	**TTO**	r 0. 86**	r 0. 78**	r 0.74 **	r 0.88 **
	VAS	r 0.98**	r 0.87 **	r 0. 93**	r 0. 80**

*Table 20 :Pearson's r values for TTO and VAS SiS and SiS+B values for Experiment 5. * Correlation significant to 0.05 level. ** Correlation significant to the 0.01 level*

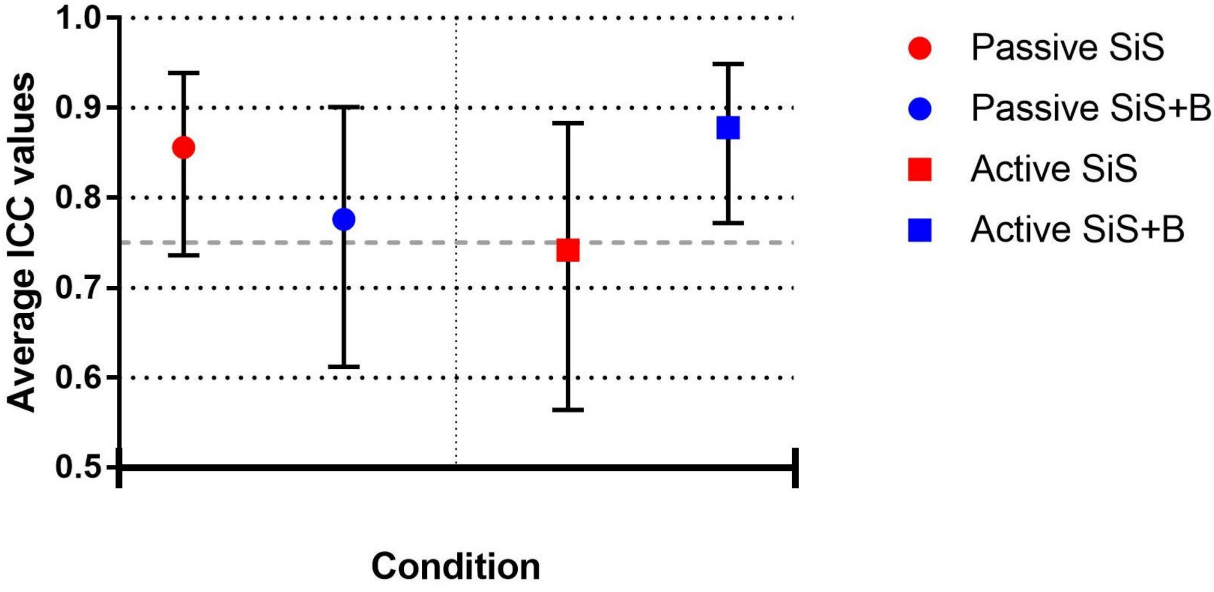


Figure 18: Mean intra-class correlation coefficients (ICCs) and their 95% confidence intervals for TTO data from Experiment 5

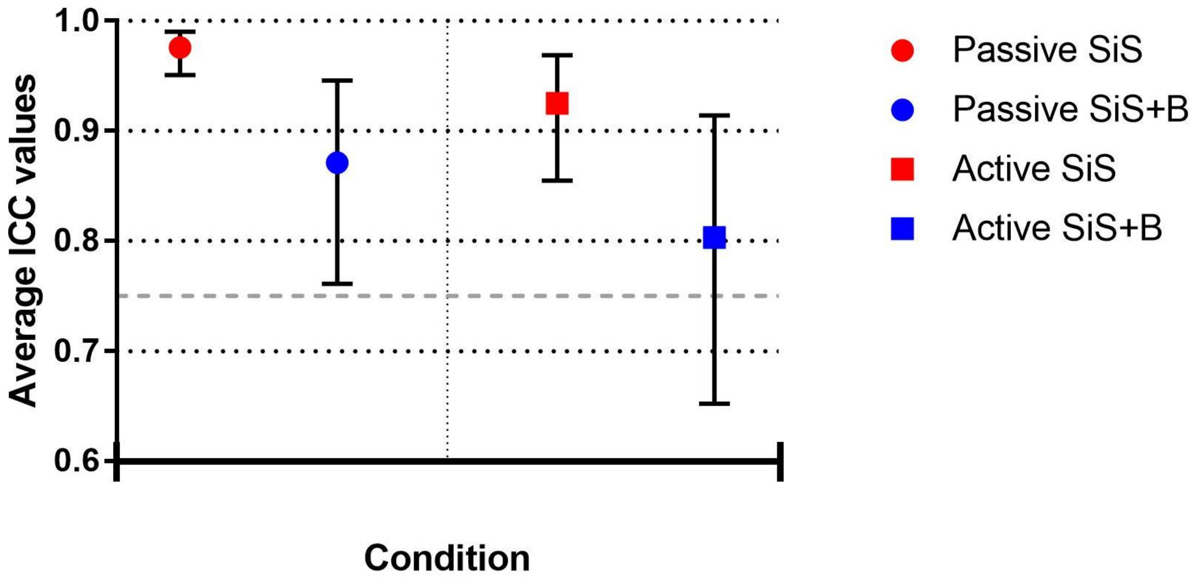


Figure 19 : Mean intra-class correlation coefficients (ICCs) and their 95% confidence intervals for VAS data from Experiment 5

5.4.4 Discussion

The reference health states collected generally indicate that the participants were in line with the population in their judgements and that the main utility task, the TTO task, was generally understood. The values for perfect health with deafness collected using the secondary utility collection task, the VAS task, was slightly higher than those observed in previous studies and the literature, however the values were not close to the figures collected for perfect health. As with experiment 4 there was a significant effect of environment and no effect of engagement observed within experiment 5. This further supports the notion that participants viewed the SiS and SiS+B health states as representing different levels of difficulty, regardless of whether active or passive engagement was facilitated. Now that the influence of an active engagement task in the health utility valuation process has been explored with in two separate experiments it is appropriate to explore and compare the effects of an engagement task when a healthy comparator is present and an engagement task when a healthy comparator is not present.

5.5 Comparing Experiment 4 and 5: The influence of presenting a healthy comparator ASH on valuations of a SSD health state.

5.5.1 Introduction

Within this chapter these experiments have explored the effects of an active or passive engagement task on utility values collected with and without an ASH of ‘healthy’ hearing to act as a comparison to the SSD hearing ASH. Experiment 4 presented two different ASH within each scenario; an example of ‘textbook’ normal hearing and an ASH of listening with SSD, experiment 5 presented one ASH within each scenario, only an ASH of SSD.

ASH are a novel tool and participants may well be hearing and interacting with these simulations for the first time, as well as encountering a health utility valuation task for the first time. It is therefore important to determine if it is necessary to present both an ASH of SSD and an ASH of healthy hearing to participants. It could be the case that by not providing two simulations; one of the target hearing state SSD and one of ‘normal’ hearing, to highlight the difference between healthy hearing and a hearing impairment the participants may miss, or potentially misinterpret, the severity of the difficulties that the simulations are portraying. Therefore, providing a reference for what ‘healthy’ hearing is could influence and impact the valuations collected. It is possible to easily answer this question by analysing the differences between TTO and VAS utility valuations collected in experiments 4 and 5.

5.5.1.1 Summary of research hypothesis

1. There will be an effect of presenting a healthy comparator on health state valuation values collected when participants are provided with an example of ‘normal’ hearing.
2. There will be an effect of presenting a healthy comparator on the reliability of health state valuation values collected when participants are provided with an example of ‘normal’ hearing.

5.5.2 Methods

5.5.2.1 Analysis

The TTO data set was automatically transformed into a utility value by the web interface, using the formula also detailed in Chapter 2 (see section 2.5.2). The VAS data were transformed into utility values (U VAS) using the formula previously detailed in Chapter 2 (see section 2.5.3). TTO and VAS values from Experiment 4 and 5 relating to research question one and two were analysed using ANOVA. The ICC values.

5.5.3 Results

The average SiS and average SiS+B TTO valuations for each participant were subjected to a repeated measures ANOVA with the within factor of environment (SiS vs SiS+B), and the between subjects factors of engagement (active vs passive) and comparator ASH (healthy comparator vs no comparator). The main effect of environment ($F(1,60)= 41.02$, $p<.00$) was significant. There was not a significant interaction between environment and engagement ($F(1,60)= .30$, $p=.59$), or environment and comparator ($F(1,60)= 25$, $p=.62$), or environment, engagement and comparator ($F(1,60)= .15$, $p=.70$).

The average SiS and average SiS+B VAS valuations for each participant were subjected to a repeated measures ANOVA with the within factor of environment

(SiS vs SiS+B), and the between subjects factors of engagement (active vs passive) and comparator ASH (healthy comparator vs no comparator). The main effect of environment ($F(1,60)= 16.08$, $p<.00$) was significant. There was a significant interaction between environment and engagement ($F(1,60)= 4.29$, $p=.04$), and there was not a significant interaction between environment and comparator ($F(1,60)= .71$, $p=.40$), or environment, engagement and comparator ($F(1,60)= .03$, $p=.871$).

ICC values have previously been calculated for repeated valuations of the same health state description within experiments 4 and 5 (SiS valuations and SiS+B valuations) (figure 22).

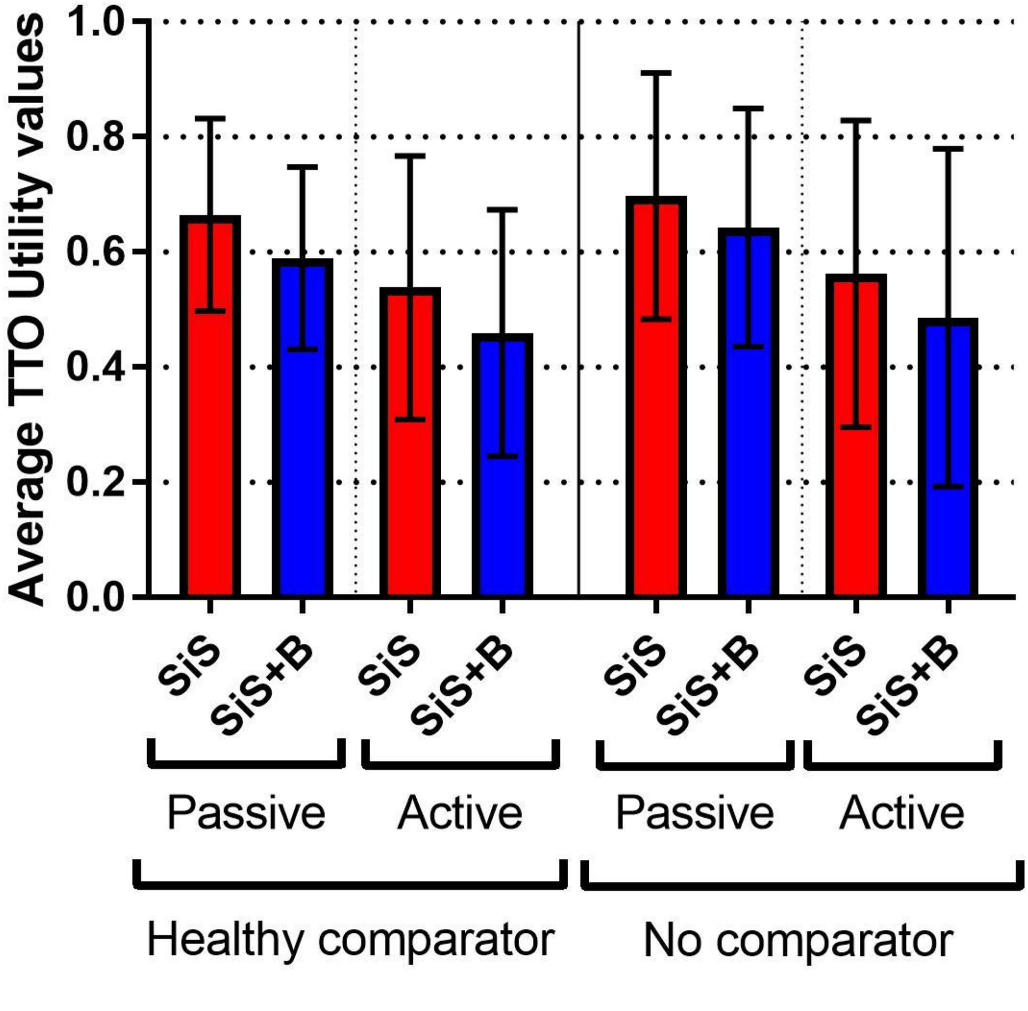


Figure 20: Average utility values for all conditions from Experiments 4 and 5. Values collected using TTO.

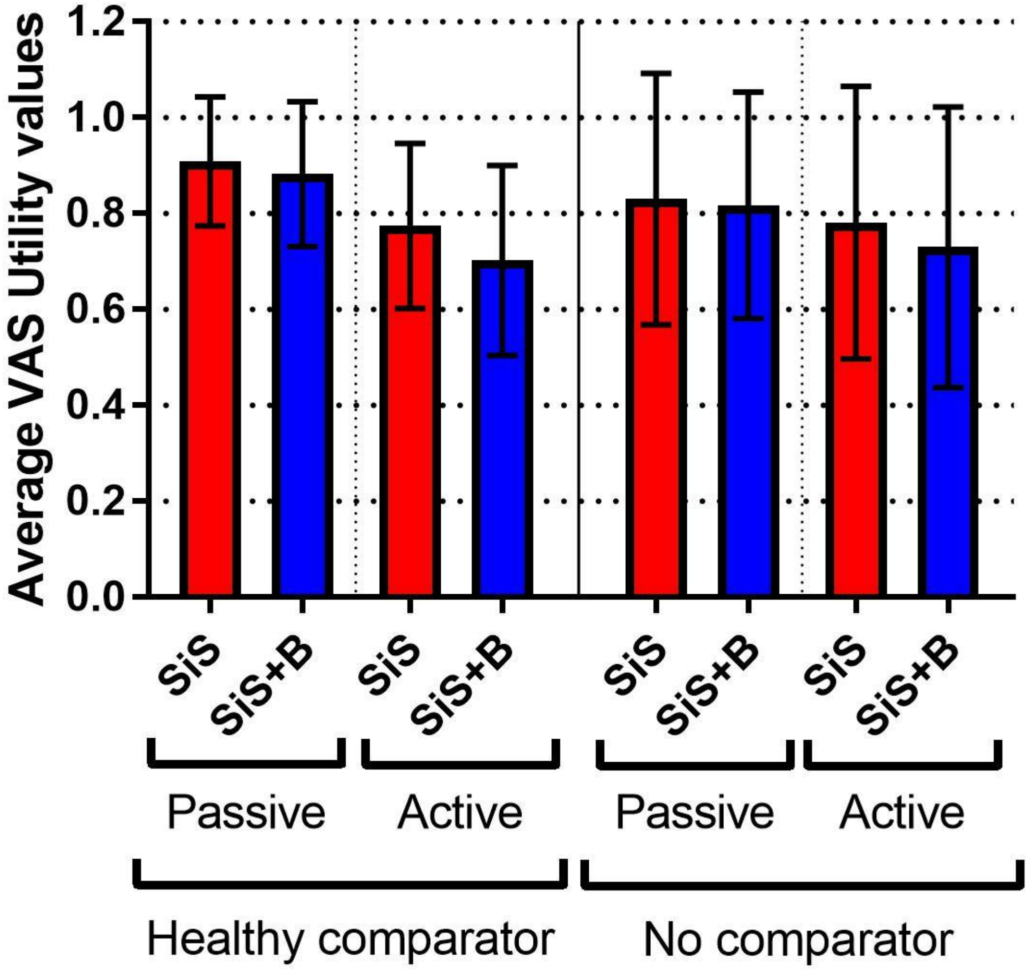


Figure 21: Average utility values for all conditions from Experiments 4 and 5. Values collected using VAS.

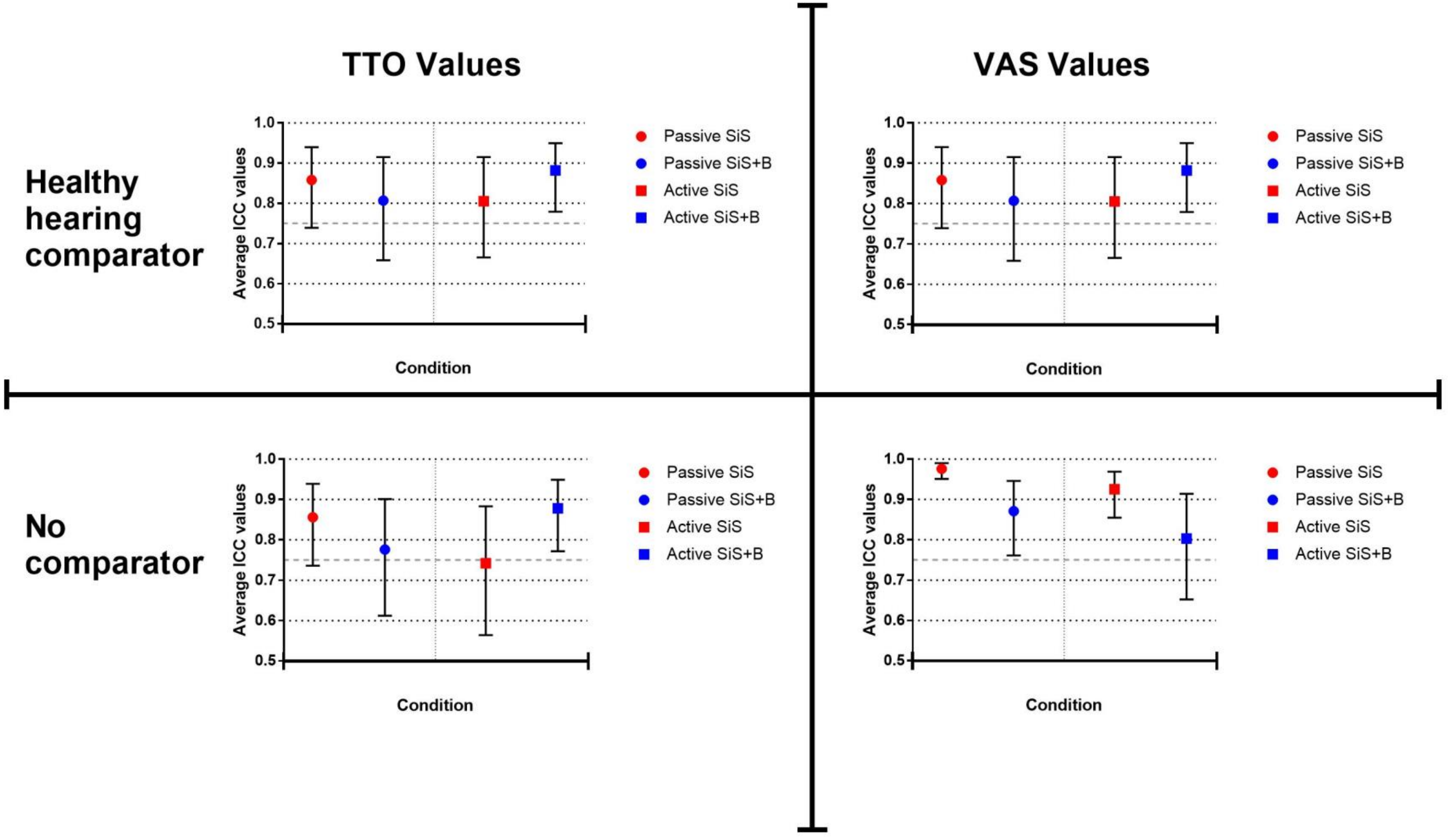


Figure 22: Layout of all ICC values calculated for all conditions from experiments 4 and 5.Values collected using TTO and VAS.

5.5.4 Discussion

When looking at valuations collected using the TTO task there was no significant effect of providing an active engagement task, nor a healthy comparator ASH on utility values, and no interaction between the conditions or factors. However, this does not necessarily mean that the addition of a listening task and the provision of a healthy comparator are useless. In fact, it could be viewed as a positive aspect that a comparator has no influence over the utility values collected within these studies. For example, if the researcher were not immediately available to provide clarification for individuals rating a health state using ASH then the inclusion of a healthy ASH could still add some level of clarification for participants, crucially without impacting nor influencing the utility values collected.

When looking at valuations collected using the VAS task there was not a significant interaction between listening environment and comparator, nor environment, engagement and comparator but there was a marginally significant interaction between listening environment and engagement, with active engagement resulting in lower average valuations (see figure 20 and 21).

It is also worth noting that when a healthy comparator ASH was presented ICC values were all above acceptability threshold for both TTO and VAS collection methods. When no healthy comparator ASH was presented there was one condition which did not meet the 0.75 threshold for high reliability, Active SiS+B. From this it is possible to infer that perhaps providing a healthy comparator ASH has a small influence on the reliability of values collected.

Therefore the results from these experiments provide further support to the idea that health valuations could be successfully provided remotely with little researcher input with an active listening task provided to reduce the risk of participant disengagement, and that the addition of this would not have a detrimental effect on any utility valuation gathered.

5.6 Experiment 6: Comparing a healthy comparator presentation with a disease only presentation

5.6.1 Introduction

When text describing any SSD health states were presented to participants in pilot 1, and experiments 1, 4 and 5, it was only a relevant section of the full text-based description of SSD that was presented. This section of text provided to participants comprised of a few sentences which pertained only to the physical hearing-related limitations present when listening to either speech in a noisy environment (SiS+B) or to a speaker during which time a separate conversation was occurring (SiS); namely an inability to localise sounds. However, normally vignettes describing the impact of health states will cover all facets of that health state, such as the physical effects of living with the condition described, the social impact of living with the condition and the emotional impact of living in the health state. Pilot 1 and experiment 1 did not use a full text description of SSD as it is not possible to simulate all aspects of the impact of a health state using ASH alone, if the full text-based description had been used then ASH may not have been directly comparable to text as they were not able to simulate aspects such as the emotional impact. For instance, the description of SSD used within this research (Lucas et al., 2018) covers the psychological impact of SSD (feelings of frustration, worrying about the hearing in the remaining ear) and the social impact of living with SSD (withdrawal from social situations), these are all aspects of living with SSD that would be incredibly difficult to simulate objectively without bias and using sound alone.

As ASH have only been used alongside excerpts of a text description, could ASH in their current state also be presented alongside a complete text-based description of SSD in order to give an individual rating the health state a greater depth of information? Presenting ASH and a full text vignette would in theory provide those rating the health state with a detailed description of the hearing-related health state and its impact, as well as a tangible experience of what it’s like to listen with a hearing condition.

Theoretically if there's a marked difference between the valuations collected using only ASH and those collected with a complete text-based description and ASH then this will be due to the depth of information now available to the participant rating the health state. This would further support the purpose and rationale behind this research project; to look into the potential of new methods which provide individuals with a more rounded and in depth description of a health state, a "bigger picture", making it easier for those with no prior knowledge of the health state to judge the impact on quality of life.

We cannot assume that the two ASH that have been developed as a part of this research, which simulate only selected aspects of what it's like to "hear" with SSD, will directly accompany and correspond with a full vignette description of SSD. Within experiment 1 there was no significant difference between valuations for SiS and SiS+B related health states presented using an ASH or a combination of ASH and a selection of the SSD text description and participants did not significantly alter the valuations they gave when the health state was presented a second time. However, this result may not be repeated when a full text description is utilised.

5.6.1.1 Summary of research questions

1. Will there be a difference between SiS+B valuations collected when an ASH is presented alongside a full text description of SSD.
2. Will there be higher reliability when the ASH and full-text description of SSD is presented to participants, rather than when the full-text description of SSD is presented using full text alone.

5.6.2 Methods

5.6.2.1 Participants

This experiment was conducted in accordance with the permissions granted by the NHS Grampian Research Ethics Committee and the Sponsor (University of Nottingham). Written informed consent was obtained for all participants by the Principal Investigator. All data were anonymised and stored in accordance with General Data Protection Regulations.

Experiment 6 was presented immediately after experiments 4 and 5 using the same participants, therefore 64 normal hearing participants (17 males, 47 females) were recruited, of these 64 participants 32 took part in experiment 4 and 32 took part in experiment 5. These participants were recruited using opportunity sampling and were a combination of participants from experiments 4 and 5. Participants had an age range of 21 to 44 years old (mean age 25.1 years) and all participants had pure-tone average threshold ≤ 30 dB HL [Hearing Level] in both ears and did not have a current ear infection or any other condition that would prohibit a hearing assessment taking place prior to the testing session. Participants were not separately financially compensated for their time completing this study as it was presented as a separate task within the study sessions for experiments 4 and 5. The experiment lasted approximately 15 minutes.

5.6.2.2 Design

Experiment 6 was a stand-alone study presented as a 'second part' to both experiments 4 and 5. This allows for a large number of participants to be recruited over a short period of time, this is beneficial as the exact power required is unknown due to the previously unused method. Experiment 6 was a brief experiment, presenting the experiment as a second part to experiments 4 and 5 also reduces duplication of effort from surplus recruitment for a brief study while recruitment occurs for experiments 4 and 5, over recruitment from the available pool of participants, and demand on available experimental resources such as sound booths.

The study was an independent measures design, the two conditions present within the study were 'full text' (FT) and 'full text and ASH' (FTA). Each participant took part in one session. In the session, the SiS+B health state was presented three times using a combination of the FT and FTA. The SiS+B simulations presented in Experiment 6 were version 3 simulations, see section 4.2.

The conditions were presented in a fixed order (Figure 23). FT was presented first, then using counterbalancing to ensure that the options were presented an equal number of times, either FT or FT+ASH was presented, then finally FT+ASH was presented. Only the SiS+B ASH was presented alongside the FT, one simulation was presented, an ASH of SSD and no listening task was given. Experiment 6 presented an ASH of the SiS+B health state alongside the FT of living with SSD. The decision to present only the SiS+B ASH was made as this health the state is the most impactful on quality of life (Lucas and Kitterick, 2014) and therefore could be more likely to indicate the usefulness of the methodology if low levels of observed power were achieved. Participants had not been presented with the full text-based description of living with SSD (see Appendix 5), although two small snippets of the FT would be familiar to participants the main bulk of the description would be new material. This lack of familiarity with the FT should result in no bias from familiarity with the description developed during participation in experiments 4 and 5.

	Presentation 1	Presentation 2	Presentation 3
Order 1	FT	FT	FT+ASH
Order 2	FT	FT+ASH	FT+ASH

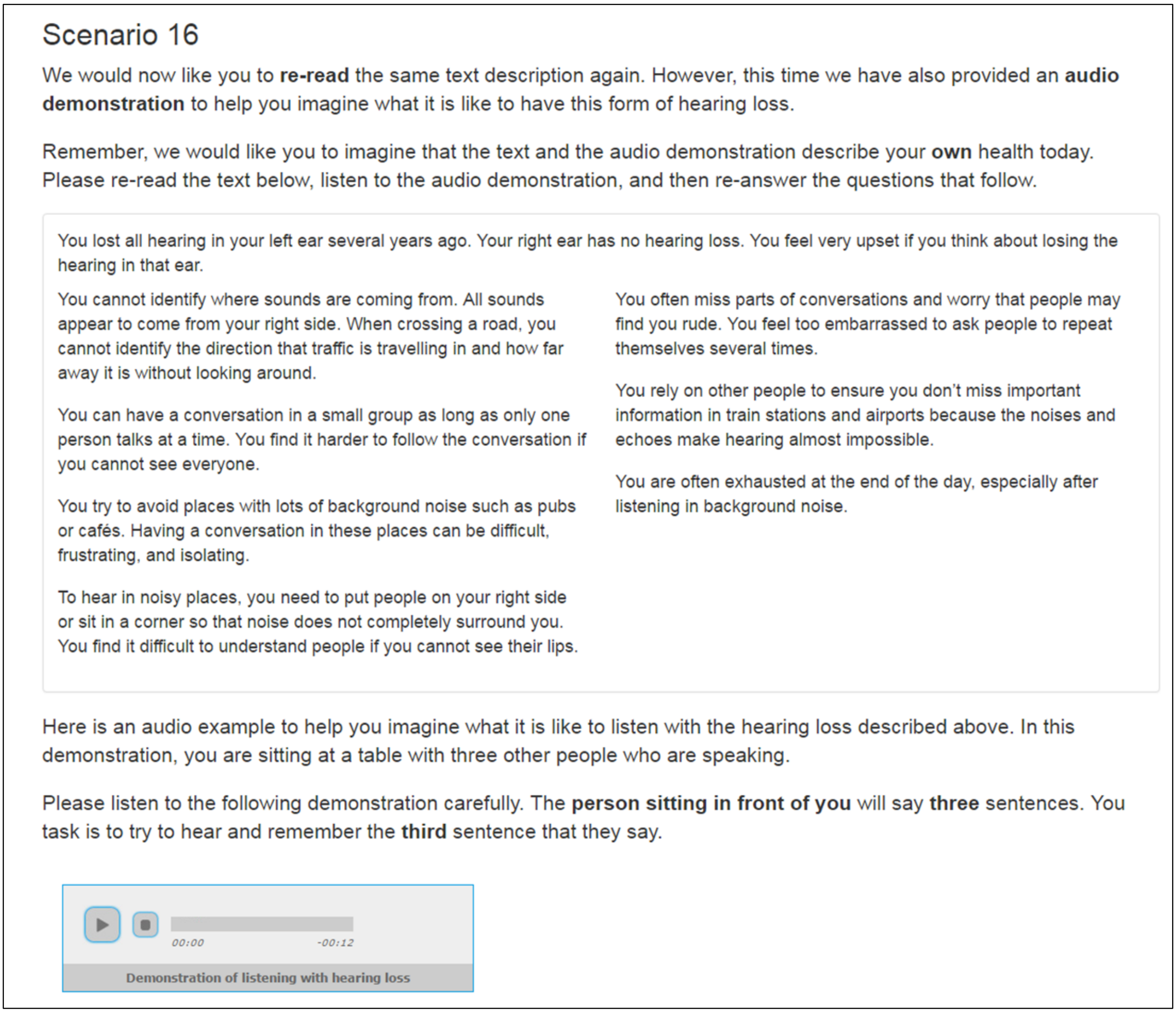
Figure 23 : Presentation orders for Experiment 6.

5.6.2.3 Sample size calculations

A post hoc power analysis was performed to ensure that adequate number of participants were recruited to ensure that statistical power was achieved.

5.6.2.4 Procedure

This experiment was presented directly after experiments 4 and 5, see sections 5.2.2.6 and 5.3.2.6 for full details of experimental set up and environment. Participants were asked to read a full text description of living with SSD and rate their quality of life as if the description represented their own health today. Participants were then asked to either: re-read the description or re-read the description and listen to an ASH of SiS+B, and then again rate their quality of life as if the description represented their own health. Finally, participants were asked to re-read the description and listen to an ASH of SiS+B and again rate their quality of life as if the description represented their own health (see figure24). Both TTO AND VAS valuation tasks were performed for each SiS and SiS+B ASH presentation.



Scenario 16

We would now like you to **re-read** the same text description again. However, this time we have also provided an **audio demonstration** to help you imagine what it is like to have this form of hearing loss.

Remember, we would like you to imagine that the text and the audio demonstration describe your **own** health today. Please re-read the text below, listen to the audio demonstration, and then re-answer the questions that follow.

You lost all hearing in your left ear several years ago. Your right ear has no hearing loss. You feel very upset if you think about losing the hearing in that ear.

You cannot identify where sounds are coming from. All sounds appear to come from your right side. When crossing a road, you cannot identify the direction that traffic is travelling in and how far away it is without looking around.

You can have a conversation in a small group as long as only one person talks at a time. You find it harder to follow the conversation if you cannot see everyone.

You try to avoid places with lots of background noise such as pubs or cafés. Having a conversation in these places can be difficult, frustrating, and isolating.

To hear in noisy places, you need to put people on your right side or sit in a corner so that noise does not completely surround you. You find it difficult to understand people if you cannot see their lips.

You often miss parts of conversations and worry that people may find you rude. You feel too embarrassed to ask people to repeat themselves several times.

You rely on other people to ensure you don't miss important information in train stations and airports because the noises and echoes make hearing almost impossible.

You are often exhausted at the end of the day, especially after listening in background noise.

Here is an audio example to help you imagine what it is like to listen with the hearing loss described above. In this demonstration, you are sitting at a table with three other people who are speaking.

Please listen to the following demonstration carefully. The **person sitting in front of you** will say **three** sentences. You task is to try to hear and remember the **third** sentence that they say.

Figure 24: screen shot of full text description

5.6.2.5 Analysis

The TTO data set was automatically transformed into a utility value by the web interface using the formula detailed in Chapter 2 (see section 2.5.2). The VAS data were transformed into utility values (U VAS) using the formula detailed in Chapter 2 (see section 2.5.3). Results from Experiment 6 relating to research question one were analysed using ANOVA. Results relating to research question two looked at the reliability of the utility values obtained and so were analysed using ICC analysis.

5.6.3 Results

The reference health state data is the same as those from experiments 4 and 5, as the data were collected after experiments 4 and 5 were completed. The reference health state data collected within experiments 4 and 5 indicated that the participants were in line with the population in their judgements and that the main utility task, the TTO task, and the secondary utility collection task, the VAS task, were generally understood.

The FT1 and FT2 TTO valuations for each participant were subjected to a repeated measures ANOVA with the factor of repetition (FT1 vs FT2). The main effect of repetition ($F(1,31) = 1.41$, $p=.25$) was not significant. The FT1 and FT2 VAS valuations for each participant were subjected to a repeated measures ANOVA with the factor of repetition (FT1 vs FT2). The main effect of repetition ($F(1, 31) = 1.21$, $p=.28.$) was not significant. The FTA1 and FTA2 TTO valuations for each participant were subjected to a repeated measures ANOVA with the factor of repetition (FT1 vs FT2). The main effect of repetition ($F(1, 31) = 4.06$, $p=.053<.$) was not significant. The FTA1 and FTA2 VAS valuations for each participant were subjected to a repeated measures ANOVA with factors of repetition (FT1 vs FT2). The main effect of repetition ($F(1,31) = 6.46$, $p<.05$) was significant.

ICC values were calculated using the value of the repeated valuations of the same health state description within the experiment (FT valuations (FT1 FT2) and FTA valuations (FTA1, FTA2)) (see Tables 22, 23, 24 and 25 for individual data sets and table 26 for associated ICC values). The ICC values for FT TTO, FT VAS, and FTA TTO were all above 0.75, and are considered stable. The ICC values for FT TTO and FT VAS data had confidence intervals that did not pass below the 0.75 threshold for acceptability (see section 3.5.2.2). The ICC value for FTA TTO had confidence intervals that passed below the 0.75 threshold for acceptability. The ICC value for FTA VAS was below 0.75 and not considered stable.

Paired samples t-test were performed on the last presentation FT data (FT1 or FT 2 dependent upon presentation order) and FTA data (FTA1) collected using TTO and VAS. For TTO data there was a significant difference between utility values collected in the FT ($M=0.58$, $SD=0.20$) and utility value collected in the FTA (M=0.47, SD=0.23) conditions, t(63)=6.46 p<0.00. For VAS data there was a significant difference between utility values collected in the FT (M=0.80, SD=0.22) and utility value collected in the FTA (M=0.75, SD=0.25) conditions, t(63)=3.41 p=0.01.

Participant	FT 1	FT 2	FTA 1
1	0.70	0.70	0.50
2	0.70	0.69	0.93
3	0.46	0.54	0.43
4	0.13	0.11	0.23
5	0.50	0.60	0.50
6	0.60	0.60	0.41
7	0.48	0.50	0.37
8	0.65	0.62	0.55
9	0.79	0.79	0.79
10	0.63	0.63	0.42
11	0.32	0.32	0.08
12	0.70	0.70	0.61
13	0.30	0.50	0.41
14	0.80	0.80	0.70
15	0.40	0.40	0.20
16	0.50	0.45	0.20
17	0.54	0.60	0.50
18	0.70	0.70	0.60
19	0.50	0.50	0.31
20	0.30	0.40	0.00
21	0.80	0.81	0.60
22	0.60	0.60	0.52
23	0.60	0.60	0.55
24	0.25	0.25	0.25
25	0.24	0.17	0.12
26	0.90	0.90	0.90
27	0.80	0.83	0.79
28	0.50	0.50	0.50
29	0.77	0.75	0.75
30	0.79	0.78	0.34
31	0.75	0.70	0.65
32	0.50	0.50	0.30
Average	**0.57**	**0.58**	**0.47**

Table 21: TTO FT FT FTA utility valuations

Participant	FT 1	FT 2	FTA 1
1	0.97	0.97	0.89
2	0.94	0.94	1.00
3	0.49	0.49	0.67
4	0.67	0.69	0.92
5	1.00	1.00	1.00
6	0.77	0.78	0.56
7	0.95	0.95	0.87
8	0.97	0.97	0.96
9	1.00	1.00	1.00
10	0.85	0.92	0.77
11	0.83	0.83	0.43
12	0.99	0.99	0.99
13	0.56	0.77	0.67
14	0.97	0.98	0.85
15	0.56	0.56	0.43
16	0.92	0.92	0.92
17	0.67	0.56	0.43
18	0.97	0.97	0.96
19	0.85	0.85	0.72
20	0.57	0.67	1.00
21	0.00	0.00	0.00
22	0.88	0.88	0.89
23	0.85	0.77	0.77
24	0.43	0.43	0.30
25	0.13	0.19	0.04
26	0.99	0.99	0.99
27	1.00	1.00	1.00
28	0.85	0.85	0.56
29	0.99	0.97	0.97
30	0.89	0.91	0.47
31	0.85	0.88	0.77
32	1.00	1.00	0.92
Average	**0.79**	**0.80**	**0.74**

Table 22: VAS data FT FT FTA

Participant	FT 1	FTA 1	FTA 2
1	0.50	0.45	0.44
2	0.50	0.31	0.20
3	0.76	0.59	0.56
4	0.65	0.53	0.66
5	0.62	0.65	0.64
6	0.64	0.48	0.34
7	0.40	0.20	0.16
8	0.74	0.60	0.51
9	0.49	0.44	0.42
10	0.55	0.61	0.42
11	0.60	0.30	0.29
12	0.50	0.60	0.60
13	0.55	0.48	0.48
14	0.60	0.50	0.50
15	0.10	0.10	0.10
16	0.40	0.30	0.20
17	0.16	0.09	0.10
18	0.72	0.62	0.00
19	1.00	0.99	0.99
20	0.30	0.41	0.42
21	0.55	0.20	0.21
22	0.60	0.50	0.45
23	0.85	0.61	0.60
24	0.74	0.79	0.61
25	0.35	0.30	0.34
26	0.65	0.16	0.30
27	0.70	0.44	0.40
28	0.94	0.97	0.80
29	0.70	0.70	0.60
30	0.77	0.64	0.68
31	0.50	0.26	0.31
32	0.10	0.10	0.11
Average	**0.57**	**0.47**	**0.42**

Table 23: TTO data FT FTA FTA

Participant	FT 1	FTA 1	FTA 2
1	0.81	0.85	0.81
2	0.85	0.77	0.67
3	0.94	0.82	0.78
4	0.95	0.90	0.96
5	0.98	0.96	0.99
6	0.92	0.58	0.54
7	0.51	0.24	0.24
8	0.85	0.77	0.70
9	0.85	0.67	0.66
10	0.92	0.41	0.30
11	0.56	0.77	0.76
12	0.56	0.67	0.66
13	0.85	0.85	0.84
14	0.97	0.92	0.93
15	0.43	0.67	0.57
16	0.85	0.81	0.67
17	0.70	0.52	0.49
18	0.97	0.90	1.00
19	0.62	0.57	0.56
20	0.76	0.57	0.72
21	1.00	1.00	1.00
22	0.81	0.77	0.72
23	1.00	0.97	0.97
24	0.95	0.99	0.90
25	0.99	0.89	0.92
26	0.97	0.85	0.97
27	0.99	0.97	0.95
28	0.99	1.00	0.98
29	0.85	0.85	0.85
30	0.85	0.68	0.75
31	0.69	0.69	0.56
32	0.16	0.16	0.16
Average	**0.82**	**0.75**	**0.74**

Table 24: VAS data FT FTA FTA

		Condition	
		FT	FTA
Valuation method	TTO	r .97 (.93 - .98)**	r .84 (.69 - .92)**
	VAS	r .98 (.96 - .99)**	r .72 (.49 - .85)**

*Table 25 :Pearson's r values for TTO and VAS SiS and SiS+B values for Experiment 5. * Correlation significant to 0.05 level. ** Correlation significant to the 0.01 level. Values within the brackets denote the lower and upper bounds.*

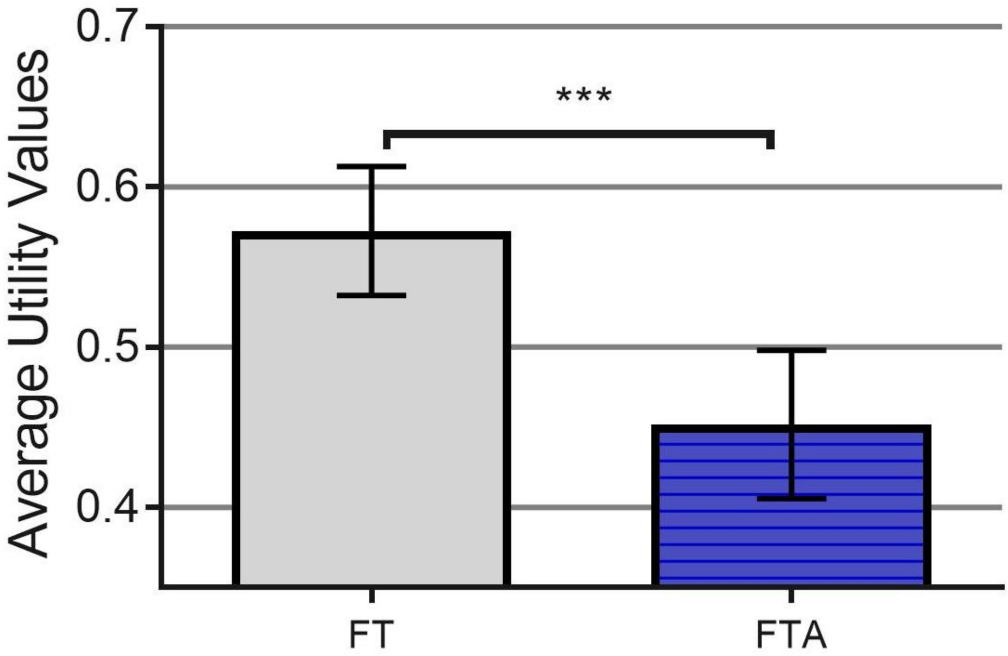


*Figure 25: Average utility valuations collected using TTO and 95% confidence intervals for all FT and FTA presentations ***denotes significant difference to the 0.01 level.*

5.6.4 Discussion

The significant difference between utility values collected using FT and FTA, indicates that there is a difference between utility valuations collected when an ASH is presented alongside a full text description of SSD. There were no significant differences between FT1 and FT2 utility valuations and FTA1 and FTA2 utility valuations respectively and the ICC values calculated show that the stability of valuations collected using both FT and FTA is high. This is not unusual or unexpected for utility values collected the full text description (the current gold standard method of describing health states), and it is encouraging to see the utility values collected using the full text description of SSD and an ASH were stable. However, it is worth bearing in mind that the stability of FTA valuations could be improved as evident from the confidence intervals for the ICC values for the FTA TTO data passing below 0.75 and The ICC value for FTA VAS being below 0.75, the level indicative of stability (see section 3.5.3).

Nevertheless, caution should be exercised before making any conclusions from this data, as it was a preliminary study to explore the suitability and usability of ASH further and SSD is described using the full text and only using an ASH of one listening situation (SiS+B).

5.7 General discussion

When looking at valuations collected using the more vigorous preference-based method of obtaining utility values, the TTO task, there is little impact of including an active listening task instead of a passive listening task to the valuation process. Nor is there any noticeable impact of providing a healthy comparator ASH to the valuation process. Valuations collected using the VAS task showed a marginally significant interaction between environment and engagement, with active engagement resulting in lower average valuations (see 5.4.4).

The lack of any clear discernible statistical difference between the type of listening task provided and the lack of influence of providing a comparator ASH could actually be advantageous practically speaking, as it could provide controls for the valuation process if it is administered in a less stringent environment, such as via an internet-based assessment tool. If the valuation task was presented using an internet based tool then the researcher may not be immediately available to provide clarification for individuals rating a health state using ASH, then the inclusion of a healthy ASH could still add some level of clarification for participants, when evaluating the impact of an SSD health state, crucially without impacting nor influencing the utility values collected. If these measures do not statistically influence the valuations collected then it appears to be safe to include both of these steps as a failsafe, to provide further clarification of the health state and to highlight the listening difficulties presented in an ASH, in case a researcher is not present to assure the valuation steps are being followed. Further support for including healthy comparator ASH is that when a healthy comparator ASH was presented reliability values were all above acceptability threshold for both TTO and VAS collection methods. When there was no healthy comparator ASH presented in experiment 5 there was one condition which did not meet the 0.75 threshold for high reliability, Active SiS+B. Perhaps providing a healthy comparator ASH has a small influence on the reliability of values collected and the inclusion of a healthy comparator ASH could increase the rigor, accuracy and consistency of this new valuation tool.

Using ASH with a complete description as a way of progressing the development of this new tool also looks promising. The collection of sensible and consistent utility valuations when SSD is presented using FTA suggests that ASH and FT can work harmoniously together, and appear to be able to complement each other, as valuations collected in experiment 6 were in line with valuations collected previously for the SiS+B SSD conditions. This indicated that, with further development work, ASH could be used to compliment a full text description as this may provide a “best of both worlds” solution; text and ASH together could provide levels of context for each other. The use of ASH could reduce and negate the incorrect contextualisation of hearing-related symptoms and their effects (such as individuals rating a text description misunderstanding unfamiliar and medically focused descriptions such as “unable to localise sounds”) while the text can effectively communicate the emotional(feelings of frustration), physical impact (tiredness), social impact (withdrawal form social situations) etc., of living with a hearing loss. It may also be possible to use a selection of ASH as a ‘bolt-on’ tool for existing QoL questionnaires such as the HUI 3 or EQ-5D, using a similar approach to Yang et al (2015).

The use of ASH as a health state description tool does not necessarily require the inclusion of a certain listening task or comparator and early indications are that an ASH can be used to compliment a FT description of SSD. There are still remaining questions surrounding ASH suitability: can ASH simulate other aspects of an SSD health state, such as treatment options in the same way it can be used to convey different levels of listening difficulty (SiS vs SIS+B)? Experiment 7 will investigate whether ASH can communicate the effect of a treatment on health utility values as text can.

Experiment 8 will explore the practicalities of administering the ASH within the valuation process using the internet. Can this new method be administered successfully to a large group of people, such as via an internet assessment tool? This would increase the amount individuals that could be consulted, and the volume of valuations collected without excessively straining existing resources.

6 Simulating interventions and the practicality of delivering acoustic simulations remotely

6.1 Introduction

Previous experiments have addressed questions about the feasibility of using ASH within the health valuation process, the validity of the simulations developed, and the optimal methods for using ASH in practice. The results of these studies have demonstrated that ASH can convey the difficulties that hearing loss can impose and different degrees of difficulty. However, in order to be useful in valuing the effects of interventions for hearing, it is necessary to use ASH to simulate treatment options. The preceding studies have also demonstrated that it is practical to integrate ASH into a health valuation exercise in the acoustically optimal environment of sound-attenuated rooms. In order to be useful in enhancing large-scale health valuation exercises, it is necessary to determine whether ASH can be delivered under less controlled conditions.

6.2 Experiment 7 – Presenting and evaluating a hearing-related health state aided by a treatment option.

6.2.1 Introduction

As shown in Experiment 1, ASH describing single-sided deafness (SSD) can be used to demonstrate what it’s like to hear with SSD in two environments that differed in terms of the level of listening difficulty they imposed. Experiments 3-6 optimised the methodology of applying ASH in practice. A logical extension of that work is to determine whether ASH can also be used to describe the effects of a hearing-related intervention on HRQoL. This step is vital in establishing the usefulness and suitability of using ASH to describe a health state, as to be able to measure the impact (and cost-effectiveness) of an intervention the impact that the intervention has on HRQoL must be measured. During the economic evaluation of an intervention, it is necessary to describe and measure the impact of both the disease in question and the impact of the treatment option being evaluated to estimate the incremental change in HRQoL.

Experiment 7 was designed to investigate preliminary questions surrounding the usefulness of ASH when describing the effects of an SSD treatment, as well as describing living with SSD. The treatment option to be simulated was a CI. A CI is highly sophisticated surgically implanted prosthesis used to restore some level of hearing to patients with hearing loss(Peterson et al., 2010, Brown and Balkany, 2007). When the delicate mechanisms within the inner ear (cochlea) that transduce movement of the tympanic membrane (ear drum) into impulses on the auditory nerve do not function correctly, then acoustic stimulation from sounds in the surrounding environment does not produce a sensation of hearing. When this happens, hearing can be restored by direct electrical stimulation of the auditory nerve, via the insertion of electrodes into the cochlea (House, 1976, Clark, 2004).

In order to restore hearing function, a CI has several distinct parts working simultaneously to pass auditory information to the brain: a microphone which picks up sounds from the environment, speech processors (both behind-the-ear and body-worn which is held onto the scalp by a small magnet implanted under the skin), a transmitting aerial, and an implantable component that comprises a receiver-stimulator and an electrode bundle that transmits sounds to the auditory nerve. The speech processors are directly involved in vocoding (a method of analysing and synthesising sounds) relevant audio information and then transmitting this data to the auditory nerve, which then passes the information on to the ascending auditory pathway and ultimately to central auditory cortices (Clarke 2004).

For Experiment 7, the effects on HRQoL of using a CI will be described using a combination of ASH and a relevant section of the text-based description (see appendices 1). SSD patients have described listening in background noise as one of the situations which they struggle with the most and has the most negative impact on their quality of life (McLeod et al., 2008). Using a difficult listening situation as the unaided comparator against which to then simulate the effects of a treatment option should in theory highlight the largest amount of benefit available from the use of a(Vincent et al., 2015). In order to recreate what it is like to hear with a CI, the ASH were altered using a vocoding algorithm (see 6.2.2.1). The ASH were not presented in isolation but instead in combination with a short text description of the health state. This choice was supported by the results of Experiments 1 and 6 which suggested that ASH can successfully be used in combination with text, and (as discussed in chapter 5) it is also most likely and practical for ASH to be used as a complementary supplement to text-based descriptions or standardised QoL questionnaires if they are eventually adopted as part of the health valuation process. The effects of the CI were also described using text alone to provide a comparison to the current ‘gold standard’ for describing health states before and after an intervention.

From the results of previous experiments, it was expected that there would be a significant difference between health state valuations collected using the two presentation methods (ASH and text vs text alone). The hypothesis being that, as with experiments 1 and 6, using the ASH alongside the text-based descriptions will add another deeper layer of detail and understanding than text alone, perhaps more successfully conveying the difficulties of the health states involved and therefore producing valuations closer to those obtained from patients directly. Furthermore, it was expected that there would be a significant difference between the health states describing the effects of cochlear implantation to the unaided SSD state. The hypothesis was that there would be a clear effect of treatment when ASH are used to describe the hearing of someone using a CI versus an unaided SSD state. It may be that the aided ASH with the CI would be seen as the more preferable health state, similar perhaps to how the SiS health state was viewed in relation to the SiS+B health state. However, no clear expectation was formed as it was unclear how participants would respond to the simulation of the CI given that it takes patients who receive a CI several months to adapt to the unusual hearing sensation it provides.

6.2.1.1 Summary of research questions

Experiment 7 aims to address the following research questions:

3. Will there be an effect of listening condition when ASH are used to describe a hearing state of someone using a CI versus an unaided hearing state?

4. Will there be a significant difference between health state valuations and the effect of describing an intervention using ASH with text versus text alone?

6.2.2 Methods

6.2.2.1 Vocoding the simulations

CI process the sounds from the real world into data that the implant can then use to stimulate the auditory nerve by using a specific signal processing strategy. The strategies used in most CI are based on vocoding analysis principles, with implants programmed with a vocoding analysis algorithm in order to turn sound waves from the real world into the binary code that the CI can then interpret. CI process sounds most like a channel vocoder (Loizou, 2006, Clark, 2004). A channel vocoder has two parts. The speech analyser divides sound up into several discrete frequency channels and extracts information about how the amplitude of the sound varies over time. The speech synthesiser takes these variations in amplitude and uses them to modulate either a series of pure tones, one for each channel, or a series of noise samples that have been filtered to match the sample frequency ranges as the channels. Finally, the modulated tones or noises are summed together to create the vocoded signal. This process demonstrates how a CI only conveys specific parts of the speech signal, albeit those that are critical for understanding speech. A channel vocoding algorithm was applied to each ASH using the computer programme Matlab (version 9.0, R2016a). The vocoder was configured with eight frequency channels across the frequency range from 350 to 8000 Hz to mimic the speech information heard by someone using a CI as these parameters reflected the approximate characteristics of most CI systems.

6.2.2.2 Approvals

This experiment was conducted in accordance with the permissions granted by the Faculty of Medicine and Health Sciences ethics committee. Written informed consent was obtained for all participants by the Principal Investigator. All data were anonymised and stored in accordance with General Data Protection Regulations.

6.2.2.3 Participants

Eighteen normal hearing participants (4 males, 14 females) with no expert knowledge of hearing research or audiology took part in the study. Participants had an age range of 18 to 40 years old (mean age 22.4 years) and all participants had pure-tone average threshold ≤ 30 dB HL [Hearing Level] in both ears from 0.5 to 4 kHz and did not have a current ear infection or any other condition that would prohibit a hearing assessment taking place prior to the testing session. Participants were recruited via an email sent within the University of Nottingham Medical School, via the public and participant engagement website ‘Call for participants’ and via placed in student areas on the University of Nottingham University park campus. Participants were financially compensated for their time at a rate of £5 per hour. The experiment lasted approximately 30 to 45 minutes.

6.2.2.4 Counterbalancing

As with previous experiments (see 4.5.2.3), to control for order effects the conditions in Experiment 7 were counterbalanced. It would have not been possible to counterbalance the order of conditions with and without ASH as it would be reasonable to assume that if a participant had first read a particular text description in combination with an ASH, then their valuation of only the text description in a subsequent condition would likely incorporate their knowledge of the simulation. Therefore, participants always valued the text alone (T-alone) before valuing the combination of text and ASH (AT). To control for order effects related to the listening condition (unaided SSD versus CI), half of the participants completed the T and AT conditions for the unaided SSD state first and then the T and AT conditions for the CI state, and half the participants completed the CI conditions first and then the unaided conditions.

6.2.2.5 Sample size calculations

Experiment 7 was powered based on data from experiment 6. The smallest effect size observed was calculated using the mean and standard deviation of a paired samples t-test on T1 and AT1 data. This resulted in an observed effect size estimate of 0.71. A power analysis for a paired t-test was conducted using the G*Power (Faul 2007, 2009) software to determine the required sample size using an alpha of 0.05, a power of 0.80, and the aforementioned calculated effect size. The required sample size was 14, and the recruitment target was 18 to account for possible attrition and possible increased variability on the data given that the effects of simulating an intervention were unknown.

6.2.2.6 Design

The study was a 2x2 factorial repeated measures design. Four conditions were defined from the combination of the two listening scenarios (unaided [SSD] and after cochlear implantation [CI]) and the two health state presentation methods (text [T-alone], and ASH and text [AT]).

The health states presented were the SiS+B unaided SSD state (see figure 26) and a CI aided state (see figure 27), presented either with text alone (T-alone) or with the combination of text and ASH (AT).

You lost all hearing in your right ear several years ago. You have no hearing loss in your left ear.

You cannot identify where sounds are coming from. All sounds appear to come from your left side.

You try to avoid places with lots of background noise such as pubs or cafés. Having a conversation in these places can be difficult, frustrating and isolating.

You are often exhausted at the end of the day, especially after listening in background noise.

Figure 26: The text description for the unaided SSD state in a SiS+B environment used in Experiment 7.

Please read the following text imagining that it describes your **own** health today.

You lost all hearing in your right ear several years ago. You have no hearing loss in your left ear. You had surgery to receive a cochlear implant (CI) which restored some hearing in your right ear.

You can tell whether a sound is on your left or right but you have difficulty identifying its exact location.

You will agree to meet in places with lots of background noise such as pubs or cafés. However, having a conversation in these places can sometimes be difficult and frustrating.

You find it easier to listen when using your CI. However, you still get tired at the end of the day, especially after listening in background noise.

Figure 27: The text description for the CI aided health state SSD in a SiS+B environment used in Experiment 7.

6.2.2.7 Procedure

As with all previous experiments participants involved in these experiments gave informed consent to participate in the study and completed a hearing screening test, which was carried out by the researcher on the day of the study. Once the consent was obtained a screening audiogram was obtained to ensure participants had normal hearing. Simulations were calibrated to within 1-dB of the 65 dB SPL target presentation level using the same procedure described in earlier experiments.

Participants were seated at a computer situated within a sound proof booth and wore a pair of closed back Beyerdynamic DT 770 Pro Studio headphones so that they comfortably enclosed their pinnae. Participants were first presented with several examples of the time trade-off task using non-hearing-related examples to familiarise themselves with the task. They were then asked to rate the reference health states from the HUI 3 (healthy, deaf, and worse state possible).

The general methodology employed in this experiment reflected the cumulative refinements developed over the course of previous experiments, including the use of an active task to engage participants in the simulations. In conditions that included an ASH, participants were first asked to listen to the simulation in its entirety and to complete a behavioural task in which they had to write down the third sentence spoken by the talker at position A. They were provided feedback on their responses by revealing what the talker actually said. Once this active task had been completed, they were then asked to rate the health state that had been described using the TTO and VAS tasks.

6.2.2.8 Analysis

The TTO data set was automatically transformed into a utility value by the web interface, using the formula detailed in Chapter 2 (see section 2.5.2). The VAS data were transformed into utility values (U VAS) using the formula previously detailed in Chapter 2 (see section 2.5.3). Results relating to research questions one and two were analysed using repeated measures ANOVA.

6.2.3 Results

The health utility results collected using TTO method are shown in Table 27, the health utility results collected using VAS method are shown in Table 28, and the average values collected using TTO and VAS are shown in Figure 28.

As with previous experiments, the average utility value obtained using TTO for perfect health was close to 1 (healthy = 0.93) (see section 2.2), the average utility for perfect health with deafness was in line with figures previously found in the literature (see section 2.XX) (deaf = 0.62), and the average utility for worst health imaginable was close to 0 (see section 2.2) (worst = 0.14). The equivalent average utility values obtained using VAS and transformed were 1.00, 0.84, and 0.23, respectively.

ANOVA were used to address the first research question, will there be an effect of listening condition when ASH are used to describe a hearing state of someone using a CI versus an unaided hearing state. The TTO valuations for each participant were subjected to an ANOVA with repeated measures factors for the listening condition (SSD vs CI) and presentation method (T vs AT), and a between-subjects factor for the order in which the listening conditions were presented (SSD then CI vs CI then SSD). The main effect of presentation was significant with AT presentation method producing lower utility values than T alone for both SSD and CI listening conditions($F(1,16)= 15.88$, $p=<.001$). The main effect of listening condition was not significant ($F(1,16)=1.18$, $p= .29$). There was a significant interaction between presentation method and listening condition ($F(1,16)= 6.04$, $p<.05$). None of the two- or three-way interactions with order were significant.

In order to understand the interaction between presentation method and listening condition, post-hoc paired-sample t tests were conducted to compare the TTO results for T alone SSD and AT SSD, T alone CI and AT CI, and AT SSD and AT CI. There was not a significant difference in the values for T alone SSD ($M=.59$, $SD=.19$) and AT SSD ($M=.56$, $SD=.23$) conditions; $t(17)=.71$, $p = .49$. There was a significant difference in the values for T alone CI ($M=.62$, $SD=.15$) and AT CI ($M=.45$, $SD=.22$) conditions; $t(17)=5.89$, $p <.00$. There was a significant difference in the values for AT SSD ($M=.56$, $SD=.23$) and AT CI ($M=.45$, $SD=.22$) conditions; $t(17)=2.22$, $p = .04$.

An ANOVA conducted on the VAS valuations produced similar results as the main effect of presentation method was significant ($F(1,14)=21.01$, $p<.001$)

but the main effect of listening condition was not ($F(1,14)= 3.55$, $p=.08$). However, there was a significant interaction between listening condition and order ($F(1,14)= 7.33$, $p=.02$), and there was a significant interaction between presentation method and order ($F(1,14)= 6.87$, $p=.02$). All other interactions were not significant.

Participant	Healthy	Deaf	Worst	T Alone CI	AT CI	T Alone SSD	AT SSD
1	0.97	0.23	0.06	0.11	0.02	0.28	0.11
2	1.00	0.60	0.00	0.60	0.75	0.76	0.65
3	1.00	0.84	0.00	0.60	0.26	0.35	0.20
4	0.80	0.50	0.10	0.60	0.70	0.70	0.60
5	0.99	0.39	0.00	0.59	0.39	0.66	0.32
6	0.99	0.89	0.08	0.83	0.81	0.85	0.72
7	0.80	0.40	0.01	0.50	0.40	0.60	0.20
8	0.99	0.50	0.01	0.70	0.60	0.70	0.70
9	0.99	0.39	0.01	0.75	0.60	0.60	0.53
10	0.97	0.96	0.21	0.40	0.30	0.70	0.41
11	0.79	0.07	0.00	0.39	0.48	0.70	0.59
12	0.73	0.75	0.44	0.73	0.80	0.75	0.80
13	0.99	0.89	0.55	0.69	0.59	0.40	0.24
14	1.00	0.25	0.00	0.41	0.50	0.59	0.29
15	0.79*	0.92	0.95*	0.99	0.70	0.60	0.50
16	1.00	0.90	0.02	0.60	0.75	0.80	0.70
17	1.00	0.90	0.00	0.50	0.95	0.50	0.20
18	1.00	0.70	0.05	0.65	0.47	0.60	0.40
Average	0.93	0.62	0.14	0.59	0.56	0.62	0.45

Table 26: TTO data for experiment 7, indicate potential anomalies within the reference health state data*

Participant	Healthy	Deaf	Worst	T Alone CI	AT CI	T Alone SSD	AT SSD
1**	0.68 **	0.93**	1.00**	1.00**	1.00**	0.98**	0.99**
2	1.00	0.85	0.00	0.95	0.96	0.99	0.95
3	1.00	0.99	0.30	0.92	0.30	0.42	0.30
4	1.00	0.92	0.43	0.85	0.97	1.00	0.85
5	1.00	0.85	0.00	0.97	0.77	0.97	0.81
6	1.00	1.00	0.16	0.96	0.96	0.98	0.86
7	1.00	0.56	0.16	0.67	0.67	0.85	0.30
8	1.00	0.92	0.43	0.85	0.85	0.92	0.85
9	1.00	0.92	0.30	0.85	0.77	0.77	0.67
10	1.00	1.00	0.57	0.85	0.77	0.95	0.85
11	1.00	0.16	0.00	0.77	0.77	0.92	0.85
12	1.00	0.89	0.30	0.92	0.92	0.99	0.95
13	1.00	0.97	0.68	0.92	0.92	0.78	0.55
14	1.00	0.56	0.16	0.57	0.66	0.77	0.56
15**	0.00**	0.99**	0.97**	0.93**	0.99**	0.98**	1.00**
16	1.00	0.97	0.16	0.92	1.00	1.00	0.92
17	1.00	0.97	0.00	0.67	0.97	0.67	0.30
18	1.00	0.85	0.08	0.87	0.77	0.85	0.77
Average	1.00	0.84	0.23	0.85	0.81	0.87	0.71

*Table 27: Transformed VAS data for experiment 7 * indicate potential anomalies within the reference health state data, ** indicates data removed from analysis.*

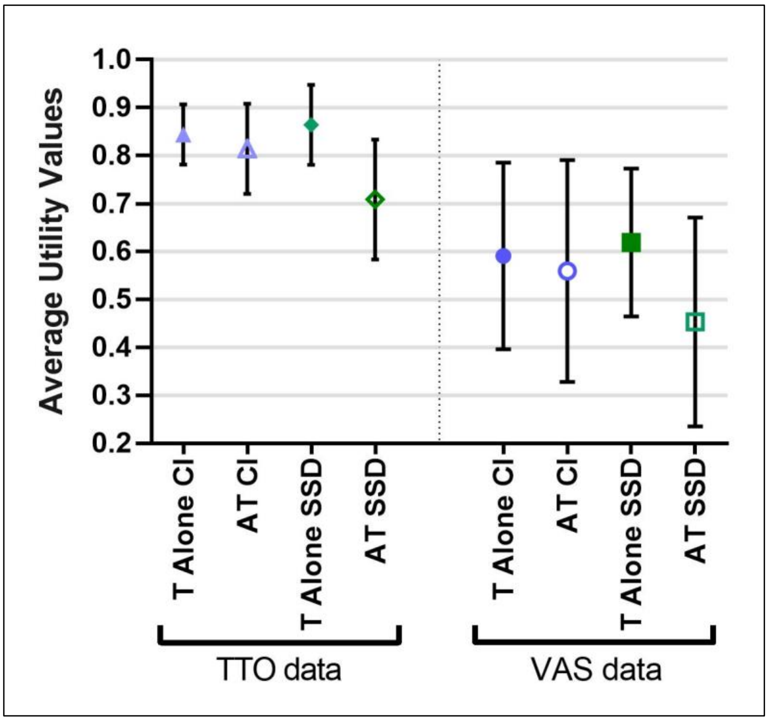


Figure 28: Average utility values and 95% confidence intervals for data collected using TTO and VAS in experiment 7

6.2.4 Discussion

The results of experiment 7 indicate several interesting findings. There did appear to be an effect of using ASH to describe the hearing health state of someone using a CI versus an SSD hearing state. Although the ANOVA performed using the TTO data showed that there was no significant effect of listening condition (SSD vs CI), the interaction term indicated that the effect of listening condition differed between the two presentation methods. When assessed for the conditions that used the combination of ASH and text, the results of a post-hoc t test indicated significant differences between the TTO valuations for SSD and CI conditions.

Anomalies were identified within the reference health state VAS data collected within this experiment. Participants 1 and 15 appear to have not fully understood the VAS task while rating perfect health, as their reference health state data were not in line with expected values. It appears these participants did not fully understand the VAS task, as the VAS values they selected for perfect health were close to zero, and the VAS values they selected for worst health imaginable were close to one hundred. These values are the opposite to what would be expected. The anomalies and potential lack of understanding of the task could have led to participants under or over valuing the health state when using the VAS technique. Two potential anomalies were present in the reference data collected using TTO, with one participant attributing a lower than expected valuation to perfect health and a higher than expected valuation to the worst health imaginable state. However, all other TTO valuations collected from that same participant appeared to be in line with expectations.

There was a significant difference between health state valuations collected using the two presentation methods, with the repeated measures ANOVAs performed on the TTO data and on the VAS data finding a significant main effect of presentation method with valuations being lower in the AT conditions than the T-alone conditions. The consistency of this effect is in line with the results of previous experiments that repeatedly found that people generally rate a state as in worse health when they have the simulation in addition to the text compared to the text alone. The significant interaction between listening condition and order for valuations collected using VAS indicated that under some circumstances the order of presentation can influence health state valuations. In this instance, utilities were generally lower when the unaided SSD states were presented before the CI states, compared to when presented in the other order. Comparisons between AT SSD and AT CI appear to indicate that there was a clear effect of listening condition (SSD vs CI) but only for the AT conditions. Therefore showing that the effects of CI using AT did have a significant effect on the utility values.

Currently, ASH have been presented to participants using a computer placed in a sound attenuated booth. Achieving the large numbers of participants required for the successful delivery of a valuation exercise but under the constraint of having to use such a research environment would be both difficult and time intensive. Delivering health valuation tasks in a less rigorously controlled environment, for example in participant’s homes by conducting the tasks online, could conceivably increase data collection rates and reduce the time demands of carrying out such work.

6.3 Experiment 8 – Practicality and validity of delivering acoustic simulations in non-controlled environments

6.3.1 Introduction

To conduct a health valuation exercise designed to assess the impact of an intervention on quality of life, researchers typically need to describe and measure the impact of both the disease in question and of the treatment in question. This process often involves gathering data from large numbers of participants which can be very time consuming and resource intensive. ASH impose additional practical requirements on this process as an electronic device and headphones are needed in order to listen to the simulations. Large numbers of the general public have access to an electronic device with internet connectivity so it is not inconceivable that ASH could be presented remotely using a website to present the valuation questions and ASH. This approach would reduce the resources required and, if the entire valuation process could be delivered online, could facilitate higher and faster rates of recruitment.

Experiment 8 aimed to assess whether it was practical and valid to present ASH and the valuation process using a website to reach a larger audience more quickly and efficiently. The experiment was a direct replication of experiment 7 but presented to participants through a purpose-built website. The hypothesis was that the pattern of results of Experiment 7, including the presence and direction of effects, would be replicated. As such, it was expected that there would be a significant difference between health state valuations collected using text alone and when ASH was provided in addition to the text description, and that the mean valuation of conditions with ASH would be lower than those based on text alone. It was also expected that a difference would arise between the health state aided with a CI compared to the unaided health state, at least when both ASH and text were used to describe those states. As participants would complete the experiment online at home without supervision, completion rates were also monitored to determine whether unattended online delivery of the experiment led to high drop out rates.

6.3.1.1 Summary of research questions

Experiment 8 aims to address the following research question:

- Can valuations be collected successfully using an online platform with minimal drop out?
- Does replicating the methodology of Experiment 7 online result in similar effects in terms of their significance and direction?

6.3.2 Methods

6.3.2.1 Approvals

This experiment was conducted in accordance with the permissions granted by the Faculty of Medicine and Health Sciences ethics committee. Consent was obtained electronically for all participants by clicking to accept the study requirements, participants were provided with instructions for withdrawal from the study should they decide they no longer wished to take part. All data were anonymised and stored in accordance with General Data Protection Regulations.

6.3.2.2 Participants

Thirty participants were recruited using opportunity sampling (16 males, 14 females), all participants had self-reported normal hearing and reported no expert knowledge of hearing research or audiology. Participants had an age range of 20 to 54 years old (mean age 29.03 years). Participants were recruited via an email sent within the University of Nottingham Medical School, and via the public and participant engagement website ‘Call for participants’. Participants were given the option to be entered into a prize draw for a chance to win one of three £10 Love to Shop vouchers. The experiment took approximately 30 to 45 minutes to complete, specific experiment duration data were not collected.

6.3.2.3 Counterbalancing

As with experiment 7 (section 6.2.2.4) it would have not been possible to counterbalance the order of conditions with and without ASH as it would be reasonable to assume that if a participant had first read a particular text description in combination with an ASH, then their valuation of only the text description in a subsequent condition would likely incorporate their knowledge of the simulation. Therefore, within this experiment, as with experiment 7, participants always valued the text alone (T-alone) before valuing the combination of text and ASH (AT). To control for order effects related to the listening condition (unaided SSD versus CI), half of the participants completed the T and AT conditions for the unaided SSD state first and then the T and AT conditions for the CI state, and half the participants completed the CI conditions first and then the unaided conditions.

6.3.2.4 Sample size calculations

Data from experiment 7 was used to power this experiment. For experiment 7, the repeated measures ANOVA had within-subjects factors of presentation method and condition, and between-subject factor of order (SSD presented first or CI presented first). There was a main effect of presentation method (AT<T alone) and an interaction between presentation method and condition (larger effect of AT vs T alone for CI than for SSD). The effect sizes (partial eta squared) were 0.498 for the main effect and 0.274 for the interaction, with the latter as the smaller effect size used to power the study. The G*Power (Faul 2007, 2009) software was used to derive the required sample size (Faul et al., 2013). A total of 14 participants would be required to conduct a repeated measures ANOVA with sufficient statistical power to detect the desired effect size at an alpha of 0.05 and a power of 0.80.

Two additional factors were taken into account when determining the final sample size. First, it was unclear whether carrying out the experiment online would cause the data collected to be more variable due to the lack of control over the research environment; e.g. the room in which the experiment was conducted, the headphones used by participants, etc. Second, it was also unclear whether the level of attrition would be higher when presenting the experiment via an online platform, as some participants may choose to abandon the session before completing the study session. Therefore, the recruitment target was set to 200% of the calculated sample size (i.e. 28 people) to account for the potential effects of these factors. As participation only required the completion of a short online series of tasks, no upper limit was placed on recruitment.

6.3.2.5 Design

The study was a 2x2 factorial between measures design. Four conditions were defined from the combination of: the two-listening scenario (monaural [SSD] and CI assisted [CI]) and the two health state presentation methods (text [T-alone], and ASH and text [AT]). Participants always valued reference health states presented using text excerpts, then valued the text alone (T-alone) presentation, before valuing the combination of text and ASH (AT). Half of the participants completed the T and AT conditions for the unaided SSD state first and then the T and AT conditions for the CI state, and half the participants completed the CI conditions first and then the unaided conditions. The simulations presented in Experiment 8 were version 3 simulations, as previously detailed in 4.2, which had then been altered using a vocoder, see section 6.2.2.1.

6.3.2.6 Procedure

Participants accessed a link to the experiment website provided via an email circular sent to university of Nottingham Medical School students or study page on callforparticipants.com. Participants were asked to complete the study when situated within a quiet environment, and to listen to the simulations using headphones.

Once participants had been given the opportunity to read through a detailed explanation of the study, its purposes, and the participant information sheet, they were then asked to give their consent to participate in the study via an online form prior to the valuation task appearing. Once consent was obtained, they were asked to use their headphones for the rest of the study and were presented with a sound check listening task (see figure 29). The purpose of this task was to confirm that participants were wearing headphones and that the headphones were delivering information audibly to both ears given that participants would need to hear the ASH in both ears to perceive the differences between the SSD and CI states.

The sound check comprised a sequence of 6 tones that were presented to both ears simultaneously, but the phase of the tones were misaligned between the ears. This misalignment induced an inter-aural timing difference so that the tones would either be perceived as being located towards the left or the right ear but only if participants were wearing headphones and could hear in both ears. The misalignment was varied randomly across the 6 tones to create a pattern of left- and right-lateralised stimuli. The task of the participant was to indicate on which side each tone was heard.

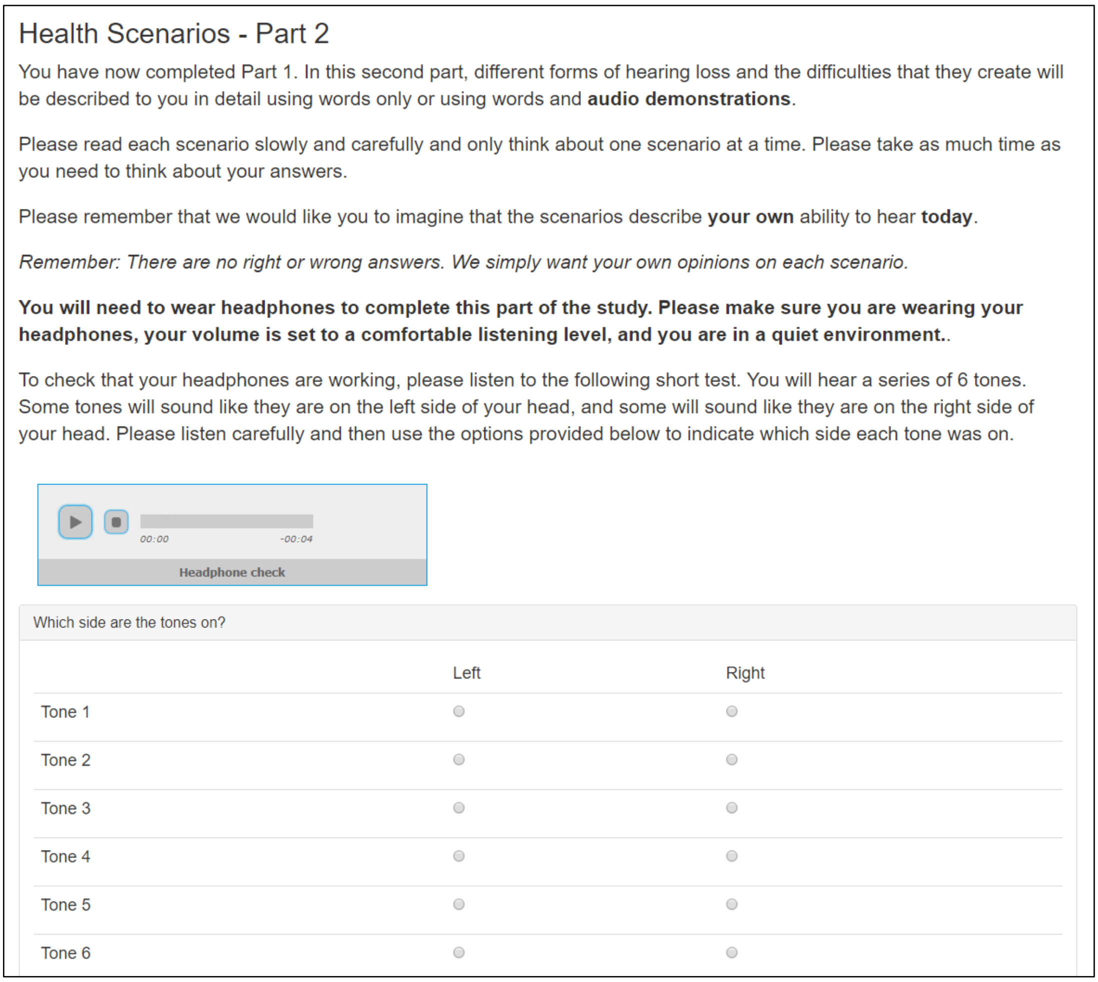


Figure 29: Sound check listening task provided to participants using the online platform

Upon completion of the sound check, participants were presented with an example of the task to come, asked to rate three reference health states derived from the HUI 3 descriptive system: perfect health, deaf but otherwise healthy derived from the HUI 3, and worst health state possible. Finally, they rated the unaided and CI health states using the TTO and VAS tasks when they were described using text alone or text and ASH. In conditions featuring an ASH, participants were asked to listen to both a simulation of healthy hearing and a simulation of health state in question. They were then asked to complete a behavioural task in which they were asked to write down the third sentence spoken by the talker at position A, and were provided with feedback by revealing what that talker actually said. Finally, they were asked to rate the health state that had been described.

6.3.2.7 Analysis

The TTO data set was automatically transformed into a utility value by the web interface, using the formula also detailed in Chapter 2 (see section 2.4.2). The VAS data were transformed into utility values (U VAS) using the formula previously detailed in Chapter 2 (see section .2.4.2). The data were analysed using ANOVA.

VAS data collected from participant 28 was removed from analysis due to anomalies in the reference health state data.

6.3.3 Results

Of the 51 participants who consented to take part in the study 30 completed the experiment, a dropout rate of 41.2%. Of the 21 participants who did not complete the experiment 9 dropped out during the collection of reference health state data and the remaining 12 dropped out during the presentation of the SSD and CI aided health states.

The health utility results collected using TTO method are shown in Table 29, the health utility results collected using VAS method are shown in Table 30, average values collected using TTO and VAS are shown in Figure 30.

As with previous studies, the average utility value for TTO perfect health was close to 1 (healthy = 0.97) (see section 2.2), the average score for perfect health with deafness was in line with figures previously found in the literature (see section 2.4) (deaf = 0.69) and average score for worst health imaginable was closest to 0 (see section 2.2) (worst = 0.11). The corresponding VAS scores were 0.93, 0.89, and 0.26. There were anomalies within the reference health state VAS data collected within this experiment: participants 3 and 13 rating perfect health very low while their other reference health state data were in line with the remaining data set. This data were included in analysis. VAS data from participants 2 and 28 were removed from analysis as the reference health state data collected were not in line with expected values, for example, participant 2 valued all health states as having high utility and consequently high quality of life, including the state of worst health imaginable which is not as expected (see 2.5.3). Conversely, participant 28 valued all reference health states as having low utility and consequently low quality of life, including the state of best health imaginable which again is not as expected (2.5.3). These results suggest that these participants may not have understood at least part of the valuation task presented to them, or they were not attending to the task, leaving the reliability of their valuations in question, for this reason their data were excluded from analysis.

The TTO valuations for each participant were subjected to a repeated measures ANOVA with factors of (SSD vs CI), presentation method (T alone vs AT), and a between-subjects factor of order (CI 1st or 2nd). The main effects of listening condition ($F(1,28)= 13.28$, $p<.001$) and presentation method ($F(1,28)=32.02$, $p<.001$) were significant. There was a significant interaction between listening condition and order ($F(1,28)= 5.53$, $p=.03$).

As with Experiment 7, in order to understand the interaction between listening condition and order, post-hoc paired-sample t tests were conducted to compare the TTO results for: T alone SSD and AT SSD, T alone CI and AT CI, and AT SSD and AT CI. There was a significant difference in the values for text alone SSD (M=.60, SD=.23) and ASH and text SSD (M=.49, SD=.25) conditions ($t(29)=2.22$, $p=03$). There was a significant difference in the values for text alone CI (M=.74, SD=.13) and ASH and text CI (M=.54, SD=.26) conditions ($t(17)=.5.13$, $p<.001$). There was not a significant difference in the values for ASH and text SSD (M=.49, SD=.25) and ASH and text CI (M=.54, SD=.26) conditions ($t(29)=-.92$, $p = .36$).

For the analysis of the VAS scores, the main effect of presentation ($F(1,14)= 21.01$ $p< .00$) was significant but the main effect of listening condition was not significant ($F(1,14)= 3.55$ $p=.08$). There was a significant interaction between listening condition and order ($F(1,14)= 7.33$, $p=. 02$) and between presentation method and order ($F(1,14)= 6.88$, $p=.02$).

Participant	Healthy	Deaf	Worst	T Alone CI	AV CI	T Alone SSD	AV SSD
1	0.99	0.30	0.00	0.40	0.29	0.29	0.50
2	1.00	0.70	0.50	0.50	0.30	0.50	0.40
3	0.94	0.59	0.02	0.50	0.78	0.65	0.77
4	1.00	0.90	0.10	0.90	0.80	0.70	0.87
5	1.00	0.56	0.03	0.60	0.15	0.25	0.35
6	1.00	0.70	0.10	0.90	0.75	0.82	0.85
7	0.90	0.89	0.01	0.69	0.49	0.80	0.70
8	0.90	0.10	0.00	0.50	0.41	0.60	0.20
9	1.00	0.71	0.06	0.74	0.58	0.74	0.72
10	1.00	0.52	0.22	0.71	0.62	0.83	0.71
11	0.96	0.89	0.11	0.80	0.50	0.68	0.13
12	0.96	0.70	0.00	0.80	0.80	0.84	0.69
13	0.96	0.98	0.26	0.71	0.06	0.43	0.42
14	0.99	0.50	0.01	0.89	0.59	0.70	0.60
15	1.00	0.99	0.21	0.94	0.79	0.81	0.01
16	1.00	0.90	0.70	0.70	0.50	0.50	0.30
17	0.83	0.59	0.30	0.61	0.46	0.70	0.45
18	0.95	0.80	0.20	0.75	0.75	0.77	0.70
19	0.93	0.67	0.08	0.83	0.62	0.82	0.71
20	1.00	0.90	0.20	0.60	0.50	0.71	0.49
21	0.99	0.21	0.01	0.70	0.02	0.01	0.45
22	1.00	0.80	0.00	0.94	0.70	0.69	0.70

23	1.00	0.79	0.00	0.81	0.93	0.83	0.84
24	1.00	0.46	0.04	0.74	0.24	0.23	0.18
25	1.00	0.80	0.10	0.90	0.60	0.60	0.80
26	0.93	0.50	0.00	0.60	0.00	0.70	0.65
27	1.00	0.80	0.05	0.83	0.70	0.76	0.00
28	1.00	1.00	0.00	0.51	0.20	0.20	0.20
29	0.96	0.74	0.08	0.84	0.75	0.42	0.36
30	0.99	0.71	0.01	0.78	0.48	0.80	0.84
Average	0.97	0.69	0.11	0.72	0.51	0.61	0.52

Table 28: TTO data for experiment 8

Participant	Healthy	Deaf	Worst	T Alone CI	AV CI	T Alone SSD	AV SSD
1	1.00	0.92	0.00	0.97	0.92	0.92	0.96
2**	1.00**	0.97**	1.00**	1.00**	1.00**	1.00**	1.00**
3	0.08*	0.91	0.07	0.65	0.92	0.92	0.94
4	1.00	1.00	0.16	1.00	0.97	0.92	0.99
5	1.00	0.77	0.30	0.77	0.23	0.30	0.37
6	1.00	0.97	0.67	1.00	0.99	1.00	0.99
7	1.00	1.00	0.16	0.92	0.85	0.92	0.92
8	1.00	0.18	0.03	0.68	0.57	0.92	0.30
9	1.00	0.92	0.03	0.90	0.73	0.86	0.92
10	1.00	0.68	0.19	0.78	0.82	0.86	0.90
11	1.00	0.98	0.16	0.92	0.57	0.85	0.31
12	1.00	1.00	0.00	0.98	0.98	0.99	0.97
13	0.07*	0.56	0.00	0.91	0.65	0.70	0.94
14	1.00	0.76	0.16	0.97	0.85	0.92	0.85
15	1.00	1.00	0.85*	1.00	0.97	0.99	1.00
16	1.00	0.97	0.92*	0.85	0.67	0.67	0.43
17	0.97	0.85	0.71*	0.91	0.93	0.98	0.93
18	1.00	0.97	0.01	1.00	1.00	1.00	0.99
19	1.00	0.85	0.30	0.96	0.65	0.98	0.64
20	1.00	0.99	0.43	0.92	0.85	0.97	0.85
21	0.92	0.77	0.00	0.43	0.01	0.00	0.55
22	1.00	0.97	0.17	0.99	0.85	0.97	0.85
23	1.00	0.96	0.03	0.99	0.98	0.96	0.96

24	1.00	0.97	0.29	0.97	0.64	0.57	0.28
25	1.00	1.00	0.30	1.00	0.92	0.92	1.00
26	1.00	0.85	0.08	0.92	1.00	0.97	0.98
27	1.00	0.99	0.13	1.00	0.97	0.99	1.00
*28***	*0.01***	*0.00***	*1.00***	*0.66***	*1.00***	*0.98***	*0.97***
29	1.00	0.92	0.09	0.95	0.95	0.75	0.63
30	1.00	1.00	0.16	0.98	0.92	0.99	0.92
Average	0.93	0.88	0.23	0.91	0.80	0.85	0.80

*Table 29: Transformed VAS data for experiment 8. * indicate potential anomalies within the reference health state data, ** indicates data removed from analysis.*

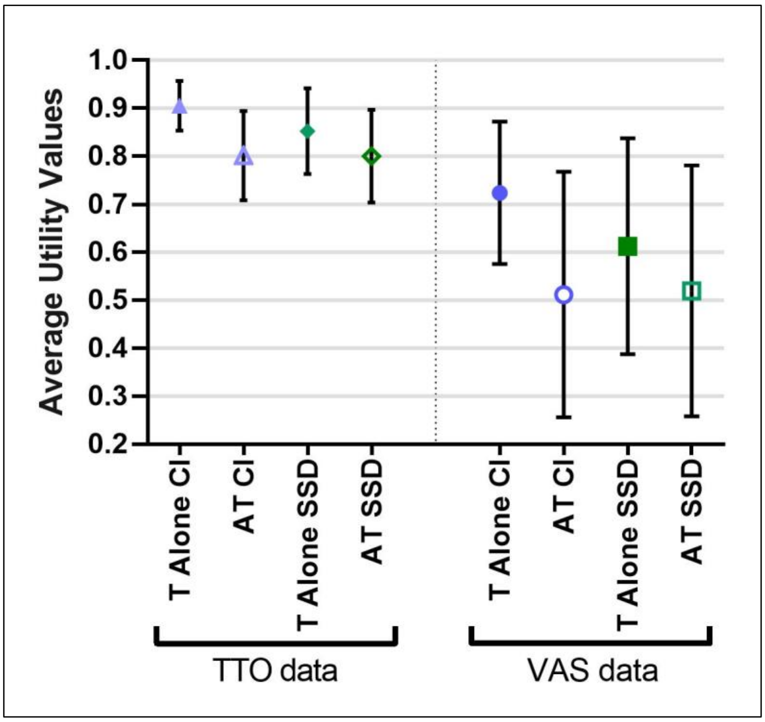


Figure 30: Average utility values and 95% confidence intervals for data collected using TTO and VAS in experiment 8

6.3.4 Discussion

Valuation data were successfully collected during this experiment, which indicates that ASH of CI can describe hearing-related health sates when presented online. However, anomalies were identified within the reference health state VAS data collected within this experiment. Participants 3 and 13 appear to have not fully understood the VAS task while rating perfect health, as their VAS values for perfect health were low, while the other reference health state data for these participants was in line with the remaining data set. VAS data collected from participant 28 were removed from analysis as the reference health state data were not in line with expected values. It appears the participant did not fully understand the VAS task when rating the reference health states as their VAS values for perfect health was zero, and worst health was one hundred. VAS data from participant 2 were also removed from the analyses as the participant gave all health states a value of one, but at the same time they appear to not be a non-trader as the participant traded off years of life when completing the TTO task.

Results indicate that there may be an effect of listening condition when ASH were used to describe a hearing state of someone using a CI versus an SSD hearing state, although the results are not clear cut. An ANOVA analysing the TTO valuations showed a main effect of listening condition, meaning that there is a difference between utility data collected using text alone and ASH and text for both SSD and CI. The ANOVA also reported a significant interaction between listening condition and order; this interaction was also found in experiment 7 VAS data. This interaction between listening condition and order for valuations collected using TTO indicates that under some circumstances the order of presentation can influence health state valuations. As with the VAS valuations from experiment 7 order one, SSD then CI, led to lower utility values than order 2, CI then SSD, and this interaction between order and presentation highlights that ASH should be presented in the same way as vignettes, with the disease-related health state presented first. Conversely, a paired samples t test looking at the TTO data for experiment 8 found that there was not a significant difference in the values for ASH and text SSD and ASH and text CI conditions. An ANOVA analysing the VAS valuations did not find an effect of listening condition but again there was a significant interaction between: listening condition and order and there was a significant interaction between presentation method and order. These interactions suggest that, as found previously within experiments 7 and 8 and within the literature, the order in which health states are presented (disease state then treatment state and visa versa) impacts on the utility values attributed to the health state. A paired sample t test also found that there was not a significant difference in the values for ASH and text SSD and ASH and text CI VAS valuations. These contrasting results across the collection methodologies could be caused by several factors. They could be due to higher variability within this data set, or due to the simulations not portraying the scenario and CI’s effects accurately. It could be that the ASH for CI could potentially benefit from subtle alterations, so that the benefits of a CI are more apparent, no changes were made to the version 3 simulations used to portray the effects of CI other than vocoding the ASH. The ASH were only vocoded due to the timeframe not allowing for a complete reworking of the simulations for CI, many hours were dedicated to building and optimising the ASH of SSD. The rationale for experiment 8 was to scope out the rudimentary ability of ASH to simulate and communicate the effects on QoL of CI usage online.

Results strongly suggest that there was a significant difference between health state valuations collected using the two presentation methods. An ANOVA performed on TTO values reported a main effect of presentation method and the paired sample t tests found a significant difference in the values for text alone SSD and ASH and text SSD and for text alone CI and ASH and text CI. An ANOVA performed using VAS valuation data reported a main effect of presentation and a significant interaction between presentation method and order. Paired sample t tests conducted on the VAS data found that there was a significant difference in the values for text alone SSD and ASH and text SSD. These results support the hypothesis that there was a significant difference between health state valuations collected using the two presentation methods, with the use of ASH and text to describe a health state resulting in lower utility valuations. However, a paired samples t test performed on VAS data also reported that there was not a significant difference in the values for text alone CI and ASH and text CI conditions. Therefore, the evidence supporting the hypothesis that there was a significant difference between health state valuations collected using the two presentation methods is strong, but not conclusive. The lack of a significant difference between the VAS values for text alone CI and ASH and text CI conditions could be due to the nature of the VAS task, it is worth noting that apparent difficulty some participants had with the VAS task led to the removal of two participants data from analysis, the removal of these data sets and the difficulties other participants could have face may have increase variability in the data.

6.4 Comparing results across experiment 7 and 8

As described previously, experiment 8 was a replication of experiment 7 that was delivered online rather than in a controlled research environment. TTO data from experiments 7 and 8 were analysed using independent t tests to assess whether there was a significant difference between data collected in the controlled research environment of experiment 7 and the uncontrolled research environment of experiment 8. There was no significant difference for the unaided SSD state described using text alone ($t(46)=-.34$, $p = .74$) or using ASH and text ($t(46)=.53$, $p = .60$). There was no significant difference for the CI state when described using ASH and text ($t(46)= -.81$, $p = .42$) but there was a significant difference between experiments for the CI state when described using text alone ($t(46)=-2.35$, $p = .02$). The difference arose because the mean valuation was lower in experiment 7 (M=0.62, SD= .15) than experiment 8 (M=0.72, SD=.15).

VAS data from experiments 7 and 8 were also analysed using independent t tests but no significant differences between experiments were identified in any of the four conditions.

6.4.1 Discussion

Comparing the results from experiment 7 and 8 it is clear that presenting ASH and an associated health valuation exercise online results in statistically comparable utility values and patterns of effects to those collected within the controlled research environment. This was the case for valuations collected using both TTO and VAS methodologies. The only exception to this finding was a significant difference in the values for the CI state when described using text alone and assessed using the TTO task. It is unclear why a significant difference arose in a condition where the health state was described using text alone given that text is the current 'gold standard' approach to describing health states. The text in question was part of a validated text description of SSD (Lucas et al 2018) and has been successfully used in previous experiments to obtain reliable utility values. It could plausibly be assumed that this significant difference could be down to preference variation within the participant cohort given the relatively small sample sizes. It is notable that very few studies have been conducted to look at the repeatability of preference judgements, or the validity of their collecting using an online platform.

6.5 General discussion

The rationale for experiments 7 and 8 was to scope out the rudimentary ability of ASH to simulate and communicate the effects on QoL of CI usage and to describe health states when the listening conditions were uncontrolled, i.e. online. These exploratory experiments indicate the versatility of ASH. They appear to be a suitable way to describe both SSD related health states and treatment-related health states. Using ASH online also initially looks to be a suitable and viable method, but larger studies that use a broader range of health states and simulations would be necessary to confirm this. The use of an online platform to present health valuation tasks could reduce the level of resource intensive work required to complete a health state valuation exercise as access to, and familiarity with, technology increases amongst the general public.

Nevertheless, the use of ASH to describe SSD related health states online is not without its pitfalls. Particular attention should be paid to the valuation methodology presented to participants. Reference health state data collected for each experiment were mainly in line with previous literature and work within this thesis, indicating ASH can indeed be used to describe health utility values remotely, from a large number of potential volunteers quickly. The evident issues some participants appear to have had with the VAS task supports the utilisation of more than one utility collection methodology to ensure viable results can be collected from a majority of participants. It is also important to note that the anomalies did not appear to occur when the TTO valuation task was used, which is an important observation as the TTO method is viewed by many as the more robust and trustworthy methodology and is a preference-based method.

ASH can be employed to describe treatments for hearing loss, such as CI, and can be presented remotely online. These conclusions are important as the methodology provides potential for a widespread roll out of a complete health valuation exercise covering the impact of both the hearing loss and of treatment options. Completing a large health valuation exercise is a very resource intensive process, taking up many research hours and with high demands on the personnel involved with the data collection. Using a remote data collection method has the potential to reduce the burden on personnel and resources. As with all research collected online, careful considerations need to be made if participants have questions or concerns about the task, as they cannot ask for help and the researcher cannot observe and step in to correct or guide a participant if the researcher believes that if they misunderstand task. There are several mechanisms that can be employed to negate these issues, such as: video tutorials or ‘walk throughs’, frequently asked question pages, and live chat functions. It may also be possible to have the participant guided through the valuation tasks in real time by a remote researcher; while this would be more resource intensive it would avoid the need for researchers to travel and help increase the number of participants recruited. All these approaches would provide participants with supplementary explanations should they require assistance when completing the valuation task, or when using the online interface.

With the main exploratory questions addressed (is the use of ASH feasible, how should ASH be presented to participants, can ASH be used to describe treatments, can ASH be presented remotely) there remains one significant question: how should the scenarios to be simulated be chosen in order to recreate what it’s like living with the condition being described. This research will now move to looking into the feasibility of creating more ASH using an existing SSD description (appendix 1) as a starting point and investigate what other scenarios, other than SiS and SiS+B, SSD patients would recommend simulating.

7 Identifying supplementary situations to simulate with ASH

7.1 Qualitative focus group study with SSD patients

7.1.1 Introduction

This chapter explores other possible situations that a group of SSD patients believe would be useful to simulate, in order to appropriately portray the difficulties of living with SSD. At present the ASH that have been developed work well, with the ASH describing two specific, discrete aspects of living with SSD; the SiS+B simulation conveys understanding what people are saying in the presence of background noise and the SiS simulation conveys listening to someone in a large group when more than one person is talking. Text-based descriptions of health states cover multiple facets of health and this includes the vignettes developed systematically by other members of researcher staff within the Cochlear Implantation team (see appendix 9.1) that have been used to describe SSD throughout this research. The development of additional ASH to present and describe other aspects of SSD and its related listening difficulties and impact could further improve ASH and their ability to describe SSD and could result in ASH communicating a more complete, multifaceted representation of what it’s like to live with SSD and how this impacts a person’s QoL. This concept is often referred to as content validity, which is an important property for a measure or test. Content validity is defined as the extent to which the content of a measurement instrument is relevant for the target population and a comprehensive representation of the construct to be measured. It is normally investigated or established through qualitative research with members of the target population (e.g. SSD patients). This type of research can also ensure that your measure is comprehensible, appropriate, and inoffensive to the target population, as they have been involved in generating the content and wording (Prinsen et al., 2016, Brod et al., 2009). To determine the most insightful and useful situations to simulate we should turn to patients and enquire about their daily lived experience and struggles, as Lucas et al (2018) did when developing the patient informed text descriptions of SSD that have been used in other experiments within this doctoral work.

7.1.1.1 Aims of the study

It is vital that any description tool appropriately portrays the difficulties of living with the health condition being described, supplementary simulations may be beneficial to appropriately portray the difficulties of living with SSD. This experiment aimed to list and itemise possible future ASH and their content, using an existing validated text-based description as a sounding board. It was not the purpose of the research to explore the impact of these situations, themes or of SSD itself.

7.1.2 Methods

7.1.2.1 Review of the literature

Thus far this research has utilised a validated text description of the effects of living with SSD developed by Lucas et al (in prep) to describe the impacts of SSD. In the Lucas et al study, eight individuals with a clinical diagnosis of SSD, who had had the diagnosis for longer than a period of 12 months and whose hearing level had been stable for the last 12 months, were interviewed to establish a text-based description of the impact of living with SSD. The participants were 4 males and 4 females, ranging in age from 37 to 71 years old. Lucas et al (2018) developed a validated description of living with SSD and within this process they reported four key themes which describe the different effects of SSD (i.e. functional consequences, psychological consequences, social consequences and positive consequences), and example situations in which individuals with SSD reported struggling (see table 31). These situations were used within this research as a foundation for a description of any new ASH to describe aspects of living with SSD.

Situation in which SSD patients reported difficulty	Number of times the difficulty was reported within the study
Identifying the direction from which sounds are coming, with and without extraneous noise	13
Understanding what others are saying in the presence of background noise	12
Inability to position themselves within a social setting so that the talker of interest is on the side of the 'good' unimpaired ear	9
Others' lack of awareness/understanding about hearing loss	9
Talking on the telephone for long periods	7
Listening to one person in a large group when more than one person is talking at a time	7
Hearing and identifying unexpected sounds (e.g. someone approaching from behind)	7
Feeling withdrawn, isolated and depressed from being unable to hear or integrate into conversations	6
Worry or anxiety about losing the hearing in the 'good' unimpaired ear	5
Unable to perceive stereo sound	5
Tiredness and fatigue, as a consequence of more effortful listening, especially in noisy listening situations	5
Ringing, buzzing or other sounds in the head (i.e. tinnitus)	5
Hearing sounds located on the side of the impaired ear	3
Dizziness or problems with balance	3
Requiring people to repeat themselves several times when unable to hear what they have said	2

Table 30 : Incidents of difficulty reported by SSD patients from Lucas et al (2018)

7.1.2.2 Participants

A sample of 10 SSD patients (4 males, 6 females) were invited to take part in the study. They were recruited through convenience sampling using the NHBRU participant data base, suitable participants were identified and sent a recruitment email. The participant database is a database of the contact details of approx. 1400 adults who are interested in hearing research. All participants were over 18, had severe SSD and had experienced their SSD diagnosis for minimum 12 months, the definition of SSD used in this study was in line with criteria defined by Vincent et al. (2015), which defined severe SSD as a profound loss in one ear (categorised by a pure-tone average threshold of above 60 dB hearing level), and normal hearing in the remaining ear (categorised by a pure-tone average threshold of 30 dB hearing level and below). Participants reported their ages within age brackets, the youngest participant fell within the ‘35-44’ age bracket and the oldest participant fell within the ‘75 and above’ age bracket. Participants were financially compensated for their time at a rate of £75 for attending the focus group, this is in line with NHBRU patient and public involvement compensation guidelines. Participants were also reimbursed for any incurred travel expenses. The focus groups lasted between 2.5 and 3 hours in total, not inclusive of comfort breaks.

7.1.2.3 Data collection methods

Qualitative research is the optimal approach for developing a rich understanding and description of an individuals’ perspectives, experiences, motivations, beliefs, and behaviours (Knudsen et al., 2012, Malterud, 2001). The specific qualitative method used in this study was the semi-structured focus group. The primary advantage of group interviews is that they promote discussion and encourage respondents to listen and respond to one other, thus eliciting shared views, contrasting opinions and novel ideas and perspectives(Knudsen et al., 2012, Lambert and Loiselle, 2008). The chief advantage of the semi-structured approach is that it facilitates the exploration of unanticipated but relevant topics, as well as the development of rapport (Leavy, 2014). Previously, focus groups have been successfully used with adults with hearing loss to develop content for video tutorials for new hearing aid users.(Ferguson et al., 2018).

This study was conducted in accordance with the permissions granted by the NHS Grampian Research Ethics Committee and the Sponsor (University of Nottingham). Written informed consent was obtained for all participants by the Principal Investigator. All data were anonymised and stored in accordance with General Data Protection Regulations. Two semi-structured focus groups were conducted, with 5 participants in each focus group. They took place in a meeting room in the NHBRU that was catered to individuals with hearing loss (e.g. loop system, good lighting, and no background noise).

Written, informed consent was obtained prior to the commencement of each focus group. All participants were interviewed by the researcher and a research assistant; each focus group lasted between approximately 120 and 180 minutes. The focus groups were audio-recorded with the permission of the participants being interviewed. After the focus groups were completed the recordings were transcribed into text files. Care was taken by the researcher to assure the respondents that they would not be identifiable in any subsequent reporting of the research findings. Data were collected using open-ended questions about situations that exemplify the main theme, these were then followed by targeted questions about the predetermined categories, such as: where does the situation occur, where are you as a listener in that situation, what makes listening difficult, who/what else is present in the environment. These questions were electronically displayed throughout the focus groups. Data were only coded and further discussion prompted on the topic if the data were directly relevant to the content of potential new ASH, and if it was feasible to simulate. This research was intended to signpost future simulations, and not repeat research investigating and detailing the impact of SSD on QoL (Lucas et al 2018). The research aimed to find and detail new ASH, therefore, any suggested situations needed to be suitable, feasible to simulate and focused on physical aspects of the effect of SSD that can be simulated using sound alone, such as difficulties localising sound, and not the emotional, or social impact such as tiredness and fatigue or anxiety about the remaining ear, as it is not possible to simulate these aspects using audio alone.

Lucas et al (2018) categorised the difficulties experienced by SSD patients into themes. In this study only themes that were feasible to simulate in the context of this research were presented to focus group participants (see table 32). It is not possible to simulate all aspects of living with SSD, particularly social impacts (embarrassment), or physical impacts (feeling tired due to increased listening effort). Participants were instructed that situations provided within the focus group sessions had to be descriptive in nature and focused on functional aspects of living with SSD. As with the simulations already developed, any new ASH would only be able to make use of acoustic information to simulate the hearing-related health states. The suitable themes detailed in table 32 were presented using a PowerPoint presentation to the participants within this experiment. These six themes were presented one by one to participants as a group in order to act as a discussion prompt.

Themes presented to participants in each focus group to prompt possible new ASH
Identifying the direction from which sounds are coming, with or without noise
Hearing sounds located on the side of the impaired ear
Understanding what other people are saying in the presence of general background noise
Listening to one person in a large group when more than one person is talking at any time
Hearing and identifying unexpected sounds
Unable to perceive ‘stereo sound’

Table 31: Key themes for potential new ASH, developed from incidents of difficulty reported by SSD patients from Lucas et al (2018)

Upon completion of the focus groups participants were asked to complete a ranking task, participants rated each of the main themes in terms of impact on QoL and importance to simulate. Impact was ranked on a seven-point scale with one being the most impactful and important to simulate and seven being the least impactful and important to simulate.

7.1.2.4 Data analysis methods

Analysis of the transcripts was supported by NVivo 12 (QSR International, Melbourne, Australia). The framework used for data analysis was content analysis. Specifically, directive content analysis was used as this is the appropriate approach for validating or extending a theoretical framework or theory. This approach uses existing research to help focus the research question and to determine the initial coding scheme (Hsieh and Shannon, 2005). The transcripts were read in their entirety and the researcher assigned each unique piece or extract of information within the transcript an individual code. A code is essentially a descriptive label.

Each code that had been generated was then assigned to a theme. A theme is a pattern of meaning or response within the data that pertains to the research question or aims(Braun and Clarke, 2006). The content analysis was deductive such that, as previously mentioned, the themes identified were based on the key concepts of Lucas et al (2018). These themes were categorised as: identifying the direction from which sounds are coming, with or without noise, hearing sounds located on the side of the impaired ear, understanding what other people are saying in the presence of general background noise, listening to one person in a large group when more than one person is talking at any time, hearing and identifying unexpected sounds and unable to perceive ‘stereo sound’ and are listed in table 33.

After an initial review of the transcripts, the codes generated by the researcher were analysed. Specifically, 30 percent of these content codes devised by the researcher were then checked for agreement by a second, independent researcher. A Cohens kappa value was calculated to determine level of agreement between the primary and secondary coders. This process is a form of peer assessment or second coding, which is a recommended technique for enhancing the rigour of qualitative research(Yardley, 2017).

7.1.3 Findings

Percentage agreement was calculated between the primary coder and the secondary coder when allocating data to themes and scenarios. The second coding was a qualitative check of whether codes were being allocated in an unbiased sensible way by the researcher. There was high agreement between the two researchers’ judgements with 97.9% agreement. The main themes and sub-themes identified from the focus groups can be found listed in table 33. The situations identified by participants are described and associated potential new ASH are detailed in the following sections.

7.1.3.1 Identifying the direction from which sounds are coming, with or without noise

Several participants suggested situations that could simulate issues with identifying the direction of sounds, both with and without background noise, such as: 1) someone calling out your name or saying “I’m here”, 2) being in a vehicle, 3) crossing a road, 4) giving presentations, 5) a crowded room i.e. a meeting or conference, 6) searching for a mobile phone, 7) smoke alarm(s) beeping.

7.1.3.1.1 Someone calling out your name or saying “I’m here”

When discussing the first suggested situation in which you enter a house and try to locate the occupants by calling out to them, participants in both focus groups mentioned issues with sound localisation. One participant reported that calling out to them is “no good to me because I just cannot, certainly in my house anyway, directionally tell where somebody is”. When discussing the localisation of sounds another participant said “it depends where they're standing when they call your name because if they are over that side [the unimpaired side] for me I could probably hear them… and I would know they were there. If they were sort of behind me here they could be calling my name I wouldn't know they were standing there. I wouldn't have heard their footsteps, I wouldn't have heard anything to tell me that there is somebody there”. Both of these quotations indicate that the direction a sound is coming from, be that from their unimpaired or impaired side, affects whether or not they will hear somebody addressing them. Participants also reported that the presence of background noise in this situation caused further listening difficulties, and that they need to be told the exact location of the person calling out. For instance, one participant said “I can hear the, “Here” but where that’s coming from and whether it’s left side, right side or behind me, I’ve no chance. So people have to say, “I’m in the kitchen” or, “I’m in the lounge” or whatever”. From these descriptions an ASH of this situation could be an individual entering a building and asking where another speaker is within that building, then being unable to locate them until the other speaker present states either their location or the room that they are in.

7.1.3.1.2 Driving in a vehicle

Several participants reported issues with hearing speech when in a vehicle, such as a car, either as the passenger or driver. One said “I’m on the right side of the car so whoever’s in the passenger seat needs to know that I can’t hear what they’re saying most of the time”. Another added “We went by car one time…he sat in the front seat…the old women were talking in the back, and we were lost. He was saying things and I said, ‘I can’t hear you, I’m sorry’, ” I couldn’t hear anything at all”, “I’m on the right side of the car so whoever’s in the passenger seat needs to know that I can’t hear what they’re saying most of the time", It was also reported that "in the car in the morning…you're trying to have a conversation and constantly having to tell them, 'You need to speak up because I can't hear you'. [The sound] needs to go up, across and down for me to pick up that." From the descriptions provided by participants, an ASH of this situation could include a back and forth conversation, possibly a speaker giving the listener directions, as indicated by a participant within these focus groups.

7.1.3.1.3 Crossing a road

Crossing a road was the third suggested situation, with participants saying that trams and vehicle horns were difficult to localise. More details about the difficulties experienced crossing a road were provided when discussing the theme 'hearing sounds located on the side of the impaired ear'.

7.1.3.1.4 Giving a presentation

Giving a presentation was mentioned as the fourth possible situation which highlighted the difficulties of identifying the direction from which sounds are coming from. An ASH of this situation could involve the listener finishing a generic presentation and then, as a participant detailed within the focus group, an audience member could ask a question, but their location within the room is not apparent when unilaterally impaired hearing is presented.

7.1.3.1.5 A crowded room i.e. a meeting

When discussing the fifth suggested situation, a crowded room in which a large meeting or conference session is occurring, a wealth of details were provided by participants. Firstly, participants said the room they would experience most difficulty in would be relatively large and possibly open plan, such as a meeting room or office, and would have hard furnishings which reverberate any sound around the room. There would be several tables within the room, seating between four to eight people at each table. Participants noted that it is more difficult to listen if these tables are round and if there is background noise and other conversations happening around this table. The ASH background noise could be the sounds of an air conditioner and/or other conversations occurring in the room. Participants suggested that a short back and forth exchange would be good to simulate, as it would be hard to follow, with a participant referencing a situation in which “more than one person [is] talking at once…want two people over there to be having their own conversation and two people over here having a different conversation”. Another participant said simulations should include “short, sharp, backwards and forwards, and have the person on the left…coming in with their contribution over the other person, because that’s normal and we do that don’t we?”.

7.1.3.1.6 Searching for a mobile phone

The sixth suggested scenario was locating a mobile phone ringing. A participant explained the potential situation as “you go from room to room to room to room. I mean it sounds silly but it can take ages. And I find that really, really an everyday situation or I'm searching for my mobile phone so I ring it from the home phone and it's going off…And you're twizzling your head. A minor, minor thing but it's something that happens every damn day”.

7.1.3.1.7 Smoke alarms

The seventh and final situation suggested as a potential simulation for identifying the direction from which sounds were coming from with and without background noise was locating a smoke alarm beeping within a room. Participants said that it is especially difficult to distinguish which smoke alarm is beeping when there were several in the environment.

7.1.3.2 Hearing sounds located on the side of the impaired ear

There were two suggested situations that could simulate issues hearing sounds located on the side of the impaired ear: 1) crossing a road, 2) someone calling out your name or saying “I’m here”.

7.1.3.2.1 Crossing a road

Crossing a road was the first suggested situation to simulate when trying to hear sound located on the side of the impaired ear. This situation was also previously suggested as a scenario which could simulate difficulties identifying the direction of a sound (see 7.1.3.1.3). Participants detailed the type of road crossing they would have difficulty with and would want a simulation to include; a central island in the middle of a busy road. One participant detailed the following situation for a potential ASH: “you’ve got a central splitter, with a bollard here and a bollard here, you're walking across… So the traffic going this way is irrelevant, you've got over, and so you're sitting there and you're going there now, and the traffic's coming that way and you can't hear it". Participants all touched upon an inability to discern where the traffic was coming from, especially if the traffic is approaching from the impaired side, with one participant recalling that when they first lost their hearing "all the traffic sounded really loud and I didn't know where it was coming from".

7.1.3.2.2 Someone calling out your name or saying "I'm here"

Someone calling out to someone was the second suggested situation to simulate when discussing the theme 'hearing sound located on the side of the impaired ear'. Again this situation was also suggested as a scenario which could simulate difficulties identifying the direction of a sound (see 7.1.3.1.1). As with the previous discussion of the situation, participants mentioned problems with the localisation of sounds and that the direction a sound is coming from. Be that their unimpaired or impaired side, affects whether they will hear somebody calling out to them. One participant reported that someone with SSD would be unable to identify where a sound was coming from "unless somebody was right there next to your ear". As with the previous mention of this situation, participants described entering a building and asking where another speaker is within that building: "this is my particular description, you come in through the door and come to the end of our hallway/entrance, the kitchen's off to the right, lounge is in front of us and the bedrooms are off to the left and, as a lot of people do, you come in the door and you go, 'I'm home' and people go, 'Yes, we're here.'

7.1.3.3 Understanding what other people are saying in the presence of general background noise

Within the focus groups there were three suggested situations that could simulate issues understanding what other people are saying in the presence of general background noise: 1) airports and train stations, 2) dining in a restaurant, and 3) a conversation with other speakers present.

7.1.3.3.1 Airports and train stations

When discussing the subject of understanding what other people are saying in the presence of general background noise, participants suggested developing an ASH in an airport and/or train station environment. Participants discussed difficulties understanding announcements, both at the station and on the train, due to general background noise, reverberation of sounds in the environment, and crowds talking. An ASH of this situation could involve an individual standing in a train station listening for an announcement about a train, as suggested by a participant “Listen for the announcements, then get on the train, make sure you get on the right one”. The listener could then get on a train and listen for the stations that it stops at “I find worst is on the train and they are saying where the stations are…if you haven’t positioned yourself to be able to see then you’ve had it because you can’t hear them and the diction’s so awful, and you’ve got the noise of the train, so that’s the worst”.

7.1.3.3.2 Dining in a restaurant

The second ASH suggested when discussing understanding what other people are saying in the presence of general background noise was a restaurant dining scenario. Participants listed difficulties with the background noise present (general background noise, music, tableware), discussions occurring on neighbouring tables, seating positions within the restaurant and talking to waiting staff. When discussing the difficulties surrounding talking to waiting staff, participants mentioned the distance between themselves and the waiting staff as being a key issue: “Then if they are trying to read back what it is you’ve ordered, they think that you’re just not interested because you can’t actually hear what they’re saying. I think that’s partly because of this distance between them just standing up and you sitting down”. A participant summarised their difficulties in such a situation: “for me, it’s just a combination of so many different things…you’re not always in control of the direction of where it’s coming from, you’ve got that difference of height that, for me, seems to cause a problem, you’ve got the background noise of people talking… It’s all that that just adds to the confusion”. Guided by this content an ASH of this situation could be an exchange between the listener and waiting staff with general restaurant background noise.

7.1.3.3.3 A conversation with other speakers present

The final situation discussed when deliberating understanding what other people are saying in the presence of general background noise was having a conversation with other speakers present. This situation was discussed briefly with a participant stating that they experienced difficulty hearing and following conversations with three or four talkers, which results in a misunderstanding of the speech. This situation was discussed further in section 7.1.3.4.2 when the focus groups were discussing the theme ‘listening to one person in a large group when more than one person is talking at any time’.

7.1.3.4 Listening to one person in a large group when more than one person is talking at any time

There were two suggested situations that could simulate issues understanding what other people are saying in the presence of general background noise: 1) a crowded room (i.e. a meeting or conference), and 2) a conversation with other speakers present.

7.1.3.4.1 A crowded room, i.e. a meeting or conference

Again, the scenario of a meeting or discussion in a crowded room was briefly detailed, with one participant stating that they were unable to locate a speaker in such situations.

7.1.3.4.2 A conversation with other speakers present

Also discussed for a second time under the theme ‘listening to one person in a large group when more than one person is talking at any time’ was a conversation with other speakers present. Again, participants noted that they experienced difficulties when conversation was scattered throughout the group of speakers and that listening to speakers present on the deaf side was difficult.

7.1.3.5 Hearing and identifying unexpected sounds

There were six suggested situations that could simulate issues hearing and identifying unexpected sounds: 1) bicycles/motorbikes passing, 2) classical music, 3) a noise from the kitchen, 4) unexpected traffic noise, 5) unexpected noises in unfamiliar environments and, 6) unexpected passing sounds.

7.1.3.5.1 Bicycles/motorbikes passing

The participants noted that both bicycles and motorbikes were difficult to hear if approaching on the impaired side and that the speed of motorbikes meant that this unexpected arrival was even more pronounced. Taking direction from a participant, an ASH could simulate a bicycle or motorbike approaching the listener from behind or their impaired side unexpectedly, in line with the description that a participant provided; “on your good side, if it’s coming up from behind you, you can hear…so you can anticipate it coming, it doesn’t matter how fast it’s coming but you anticipate it coming. The difficulty is, if it comes past your deaf side from behind…there’s no visual there for you, and by the time you pick it up it’s a lot closer to you”.

7.1.3.5.2 Classical music

The participants also discussed listening to classical music. The specific difficulties noted were with drums, or sudden loud noises within a piece of music as they can be unrecognisable and distorted. A simulation of this could be simple to develop, as a piece of classical music recorded in stereo could be selected and then played monaurally.

7.1.3.5.3 A noise from the kitchen

Hearing a noise from the kitchen was suggested by a participant when discussing hearing and identifying unexpected sound, in particular a participant briefly referenced somebody shouting out or something falling to the floor in the kitchen.

7.1.3.5.4 Unexpected traffic noise

Unexpected road vehicles were also suggested as a potential ASH to describe hearing and identifying unexpected sounds with SSD. Participants mentioned passing lorries and emergency vehicle sirens as sources of difficulty both when walking down the street and/or when driving a car.

7.1.3.5.5 Unexpected noises in unfamiliar environments

Unexpected noises in unfamiliar environments were briefly discussed by participants as a potential ASH to describe hearing and identifying unexpected sounds with SSD. Participants clarified that unexpected noises in a familiar environment included being in a new house and hearing the noises of neighbours, such as chairs moving, or plugs being put into sockets in the neighbouring property.

7.1.3.5.6 Unexpected passing sounds

Finally, briefly mentioned by a participant when discussing and identifying unexpected sounds was passing sounds, in particular one participant suggested a passing jogger or simply someone walking by.

7.1.3.6 Unable to perceive ‘stereo sound’

Within the focus groups there were two situations suggested when discussing the theme ‘unable to perceive ‘stereo sound’’: 1) listening to the television, and 2) listening to stereo music.

7.1.3.6.1 Listening to the television

The first situation mentioned when discussing the inabilities to perceive ‘stereo sound’ was listening to the television. Participants mentioned experiencing particular difficulties if there are multiple voices present within a television program, as they will think somebody in the room has spoken, or if somebody in the room does speak then their speech is difficult to hear. A suggested situation from a participant was “listening to a television programme and when there’s a dialogue going on, get somebody to speak in the room”.

7.1.3.6.2 Listening to stereo music

The second situation briefly detailed when discussing the inabilities to perceive ‘stereo sound’ was listening to stereo music. Participants discussed losing sound quality and that, as music is produced in stereo, monaural listeners lose the stereo aspect, as SSD listeners can only hear half of the music presented. A simulation of this situation could involve presenting a piece of stereo music monaurally (i.e. through headphones but with the sound sent to one ear presented at a much lower volume), the piece of music could be one suggested by participants, such as Bohemian Rhapsody by the band Queen, which is recorded in stereo and presents aspects of the music differently to each ear or less well-known music or a radio recording “where the interviewer is in one speaker and the respondent is in another, with just a small element of transfer between them”.

7.1.3.7 Miscellaneous

There were several noteworthy mentions of environments that lead to listening difficulties which were not expressly connected to the main themes, but which could be noted and evaluated as potential aspects to include in simulations, should more ASH be produced. Participants stated that softly spoken individuals, general outdoor noise and environments that echo, such as swimming pools, were situations which caused difficulty. Participants also mentioned difficulties with unfamiliar voices, such as talking to strangers.

Main theme	Situation	Situation as described by participants
Identifying the direction from which sounds are coming, with or without noise	Someone calling out your name or saying "I'm here"	""Where are you?" and somebody goes, "Here""
		"That is no good to me" because I just cannot, certainly in my house anyway, directionally tell where somebody is"
		"I live in a two storey house and there's a noise. Yeah or the husband, will say. Well I'm here"
		"And you go from room to room to room to room"
		"when I was a teenager and I'd maybe be walking through the centre of town and I hear somebody shout my name. I stood back looking around and they're just shouting my name"
		"it depends where they're standing when they call you name because if they are over that side for me I could probably hear them and I would know they were there. If they were sort of behind me here they could be calling my name. I wouldn't know they were standing there. I wouldn't have heard their footsteps I wouldn't have heard anything to tell me that there is somebody there even that can be one hell of a shock"
		"when we come back from work you know just trying to work where they are. Are you upstairs are you downstairs. Are you at the back of the house"
		" if there's background noise it's worse for me"
		"I can hear the, "Here" but where that's coming from and whether it's left side, right side or behind me, I've no chance. So people have to say, "I'm in the kitchen" or, "I'm in the lounge" or whatever"

		"They're not telling me where they are. I don't know where to look"
		"If somebody is shouting at you. You need them to shout repeatedly so that you can turn and wait until it gets louder and and then find where it's loudest and then go that way"
	Being in a vehicle	"I'm on the right side of the car so whoever's in the passenger seat needs to know that I can't hear what they're saying most of the time"
		"So the siren goes, and until I locate it I pull over and stop, and I'm looking for the lights. I find that quite dangerous"
		"We went by car one time, and exactly as you described, he sat in the front seat, the girls were talking- Girls, the old women were talking in the back, and we were lost. He was saying things and I said, "I can't hear you, I'm sorry." I couldn't hear anything at all"
		"I'm on the right side of the car so whoever's in the passenger seat needs to know that I can't hear what they're saying most of the time"
		"if they're strangers that makes it socially awkward"
		"in the car in the morning when you're trying to have a conversation and constantly having to tell them, "You need to speak up because I can't hear you. It needs to go up, across and down for me to pick up that.""
		"Crossing roads "

	Crossing a road	"waiting for a tram, and like kids of that age did, and probably still do, I used to smoke. I took out a cigarette and I dropped it, it rolled into the tram tracks, I bent down to get it and as I picked it out a tram whizzed past my head"
		"Crossing the road. And if a car or a lorry pipping its horn"
		"everybody was screaming and I couldn't hear it because it was on this side"
	Presentations	"giving a presentation and someone asks a question and they don't actually indicate who they are then I don't know who they are. I look around and say, "Who said that?""
	A crowded room i.e. a meeting or conference	"In a relatively crowded situation"
		"a room of about this size with about 30 or 40 people in discussing cases"
		"in a work situation I agree with [Name], the worst is when they break you out into this, they say, "Go into twos or threes" in the same room"
		"conferences and that sort of thing when you sit at a round table it's horrible"
		"they're trying to do in that icebreaker is get you to talk to people and everybody's trying to talk. So you've got all these people all trying to talk about, "What did you eat last week for breakfast?", or whatever it is, and for me that's a really difficult situation to be in, because of what we've discussed. Everybody's trying to talk, and the whole point of that exercise is to try and get people to understand a bit more about everybody else around that table"
		"Background noise and other people"

		"the office environment"
		"everyone crowded round the bed and I used to have to get as near as I could and I just couldn't hear what on earth was going on at the bedside"
		"and there's someone talking in the corner then the air conditioning's in the way and I can't hear"
		"a conference situation where you've got lots of discussions happening in the same room"
		"Say we all broke out into little groups and you had you two talking and you two talking, then I wouldn't be able to hear"
		"Yes, lots of different conversations"
		"that problem of other people talking and the noises going on"
		"The type of room as well, because the sort of place where you'd have a conference has got very little soft furnishings, so the reverberations are enormous and it multiples the difficulty"
		"Can we go there because they've got soft furnishings and definitely not there because it's a solid floor"
		"Hard services are a good one to demonstrate the problem"
		"a round table, as [Name] just said, is a specific problem"
		"circular tables I think are the ultimate problem because you can't choose your best position"
		"Lots of other tables in the room, all round tables"
		"the worst place for me would be in that middle table. If you've got tables dotted around a room"

		"Participant 1: Six or eight people on them. Participant 4: Four, five, six… This is a very standard conference situation"
		"a break out session with them all talking at the same time"
		"It's the peripheral sounds and having everybody at your table talking at the same time"
		"Have some noisy aircon on or maybe some music"
		"Or background music"
		"A bigger room"
		"a solid floor rather than a carpeted floor you get more bounce off that"
		"no damping for the sound, all the people are there chattering away and it seemed"
		"it's got a hard floor and there are no curtains to soften it up"
		"A lot of glass"
		"the person who is you sitting there and make sure the person to whichever side you're going to make deaf, let's say it's left for now, make sure the person on your left is joining in the conversation and contributing because when you take out the left side they're going to be non-existent"
		"more than one person talking at once. So you want two people over there to be having their own conversation and two people over here having a different conversation"

		“short, sharp, backwards and forwards, and have the person on the left, assuming it is the left, coming in with their contribution over the other person, because that’s normal and we do do that don’t we? “We used to-” “Did you go to so-and-so?” or whatever it might be”
		“I was in a big meeting on Tuesday. I know most of the people who are going to be in that meeting. But just concentrate on direction which sounds are coming from somebody called out my name. Um big room lots of noise going on but I could hear it”
		“environment I work in and you know people have come up to me yesterday and today. There's just too much background noise and they're even right there on my good side. But there's two people just directly behind them having a conversation. There's phones going. There's photocopiers. The Doors are going and it's just you know I've got to really really struggle. Really strain to hear what they're saying”
		“Open plan”
		“The more background noise the worse it's going to be”
		“air conditioning. It's just so loud”
		“rooms with no soft furnishings if it's echoey and very large open spaces and if they are like a children's play area there's lots of noise going on”
		“they don’t actually indicate who they are then I don’t know who they are. I look around and say, “Who said that?””

		“The side will get scrambled, so I might hear a bit of what [Name] saying over there, a bit of what you’re saying, and it all gets mixed up into the same conversation”
		“at the round table, it doesn’t matter where you sit, you’ve always got somebody there, and it’s embarrassing”
		“if it’s chat conversation, “Let’s get to know each other”, that makes it a lot harder because you’re getting short elemental bits, “I lived in Derby, I went to this…” and you’re not getting anything, then it’s back to you to say your bit, and it’s going backwards and forwards. You’re missing it because the fella next to him is saying something to somebody else”
		“You get bits of this other conversation going on and you can’t distinguish it from this conversation, so you can’t tune into the bits you want to tune into”
		“Sometimes. It becomes a soup. So instead you can't pick out the individual ingredients”
		“I find I can manage the odd word or sentence out but then I’m like suddenly oo I missed that one”
	Searching for a mobile phone	“you go from room to room to room to room. I mean it sounds silly but it can take ages. And I find that really really an everyday situation or I'm searching for my mobile phone so I ring it from the home phone and it's going off. But you know. And you're twizzling your head. A minor minor thing but it's something that happens every damn day”

	Smoke alarms beeping	“the smoke alarm when you've got loads of smoke alarms in the house. And one peeps because the battery has gone”
		“it’s so difficult to tell because they're not that far away from each other although there is one downstairs and one upstairs”
Hearing sounds located on the side of the impaired ear	Crossing a road	“the danger, as [Name] described, is the crossing of roads”
		“all the traffic sounded really loud and I didn’t know where it was coming from”
		“You could have a very sudden- You could have quiet and then when the car is there you suddenly notice it so it really makes you jump”
		“You’ve got someone shouting at you like you nearly jumped in front of a tram. People do that on my good side, people are forever pushing me back”
		“a central splitter island in the middle of the road, where you’ve crossed half the road and you’re now waiting to cross the second half of the road”
		“a road like this, and you’ve got a central splitter, with a bollard here and a bollard here, you’re walking across there… So the traffic going this way is irrelevant, you’ve got over, and so you’re sitting there and you’re going there now, and the traffic’s coming that way and you can’t hear it. You can’t hear it as well, you do hear something but you’re not quite sure where it is, it’s not as well established as is on that side”

		"if you're stood there and you're asking the question, "When do you cross?" if you're hearing this constant stream behind you you're sitting there or standing there thinking, "Do I really go? Do I really go?" Then it comes back to what [Name] was saying, you then have to rely visually on seeing whether something is there or not"
		"Interviewer 1: Is there a particular type of vehicle that would best portray that? Participant 2: A motorbike"
		"have almost gone under a bus because I looked the wrong way"
		"your normal instinct tells you, "Look in this direction" and you almost think, "I've done my good bit, I've done what I've been taught in the green cross code", and all the rest of it, but you don't naturally look to the other side"
		"with the car, I don't think you'd pick up direction"
		"you haven't got a clue anywhere where it's coming from, unless you wait until someone comes along and walk with them"
		"You're waiting for the vehicles coming this way, and without the hearing in your left side you're not going to have as good a chance"
		"If it was coming at you from your good side you make that connection and go, "Okay, it's still quite a way away" and you hear it coming towards you so you almost gauge it. But when it comes from your not good side because I can't gauge how far away it is when I have heard it I'm

		automatically just a bit more nervous about, “I’ve heard something and I need to be more on my guard.””
		“on your good side you hear the sound directly, on your bad side you hear a reflected sound, so depending on the environment you hear that reflection later and later”
	Someone calling out your name or saying “I’m here”	“this directional one, having people saying, “I’m here””
		“in my house I have to stand there and go, “Tell me where you are””
		“this is my particular description, you come in through the door and come to the end of our hallway/entrance, the kitchen’s off to the right, lounge is in front of us and the bedrooms are off to the left and, as a lot of people do, you come in the door and you go, “I’m home” and people go, “Yes, we’re here.””
		“you can’t see them so in scenario you’re totally reliant upon what you hear. You can definitely hear somebody saying something but you haven’t a clue where that person is”
		“We’d not know where the sound was coming from unless somebody was right there next to your ear.”
Understandin g what other people are saying in the	Airports and train stations	“airports or train stations, you can’t hear the announcements”
		“Waiting at the train station”
		“there is a load of background noise”
		“The airport announcements, the railway station announcements, I hear some but not all”

presence of general background noise		“I find worst is on the train and they are saying where the stations are, or whatever, and if you haven’t positioned yourself to be able to see then you’ve had it because you can’t hear them and the diction’s so awful, and you’ve got the noise of the train, so that’s the worst”
		“a noisy train”
		“Listen for the announcements, then get on the train, make sure you get on the right one”
		”there were people around you talking”
		“a crowded train”
	Dining in a restaurant	“out for a meal with your family, you’re in the restaurant”
		“there’s lots of noise going all round, and, rightly or wrongly, I can’t necessarily hear the waiter or the waitress very well”
		“if they’re slightly mumbling or there’s all this background noise I find it difficult to pick up”
		“Then if they are trying to read back what it is you’ve ordered they think that you’re just not interested because you can’t actually hear what they’re saying. I think that’s partly because of this distance between them just standing up and you sitting down”
		“where there are different people having their own conversations”
		“you’ve got mealtimes and other tables having their discussion”
		“General restaurant noise”
		“Music on in the background”

		“if the waiter appears higher and on the other [deaf] side”
		“The more people the harder it is”
		“the group don’t generally subside when you’re talking to the waiter, they’re doing their own conversation”
		“it’s worse in large ones if there are lots of other people, yes”
		“there’s more people and there’s more background noise, there are more plates clanking”
		“for me, is just a combination of so many different things. One is that you’re not always in control of the direction of where it’s coming from, you’ve got that difference of height that, for me, seems to cause a problem, you’ve got the background noise of people talking, and they’ll not even be talking about the same thing they’ll be having general conversations like people do when they go out to eat. Some people will be talking about schools and talking about… It’s all that that just adds to the confusion”
	A conversation with other speakers present	“Going to the understanding part of that sentence, one of the things that I find and I don’t know if anyone else does, if I’m in a general conversation with three or four people talking there are two scenarios, one is I bring up something that is totally dissociated from the conversation because I’ve misunderstood what they’ve said. But then the other one is when I repeat word for word almost what has just been said by the person on my left”
Listening to one person in	A crowded room, i.e. a	“if someone’s talking in a big group and you’re looking round to see who’s talking because you don’t know who’s talking you miss what they’re saying. If they’re saying something to you that’s

a large group when more than one person is talking at any time	meeting or conference	important for the meeting, say if you're chairing it or something like that, it can be very difficult. You can't see who's talking, then you suddenly locate them, by which time they've finished talking because they've made their contribution and you've completely missed it"
	A conversation with other speakers talking	"a group of people not like one main speaker and then other people talking"
		"you'd also not hear so if some are almost sort of towards your bad [deaf] side. So you're only getting part of the conversation"
		"when the conversations you know throughout the group rather than one conversation"
Hearing and identifying unexpected sounds	Bicycles/motorbikes passing	"When a bike comes up that [deaf] side it's terrible isn't it, a bicycle? Bicycles are so dangerous, they shoot up like this and if they're on this [deaf] side not only have you not heard them but they do that thing and appear from nowhere"
		"It's all quiet and then suddenly something zooms past you"
		"motorbikes, because even with the best will in the world they always tend to go rather quickly and I don't know if it's to do with their size as well the bike"
		"Interviewer 1: Does it make any difference to your awareness of the motorbike whether it's coming up behind you or whether it's coming towards your deaf side? Participant 2: Behind. Participant 4: I think it's worse behind. Participant 3: Yes.

		Participant 4: On the deaf side. Participant 1: Yes, because then you don't hear."
		"on your good side, if it's coming up from behind you, you can hear it coming up from behind you so you can anticipate it coming, it doesn't matter how fast it's coming but you anticipate it coming. The difficulty is if it comes past your deaf side from behind you can't see it, there's no visual there for you, and by the time you pick it up it's a lot closer to you"
	Classical music	"When there was any sudden drums or other sudden noise. The sound was so distorted that was the thing it wasn't like the purer it was like. You know I was jumping out of my seat like god what was that."
		"And it could be an everyday noise but it is totally totally. You know. Well it's just unidentifiable to what it is"
		"Something loud in a piece of music"
	A noise from the kitchen	"an unexpected sound in the kitchen"
		"somebody shouts or something lands on the floor in the kitchen"
	Unexpected traffic noise	"A lorry just coming up"
		"A police siren going or an ambulance"
		"It's an unexpected sound and they're going to go fast and if you are going to walk over to them they're not going to stop and a car will beep its horn. And hopefully stop. But it is all of a sudden it is an unexpected sound. You know you're just not listening to normal traffic and stuff like that.

		So it's an unexpected sound that's the difficulty. that's the problem, something coming behind you”
	Unexpected noises in unfamiliar environments	“being in a new house with sounds coming from neighbours”
		“somebody next door. Plugging something into the wall and if you hear it through the wall what that, where's that come from?”
		“next door neighbour pulling his chair out across the floor and it scrapes across the floor.”
		“banging on the walls”
	Unexpected passing sounds	“somebody might just walk past or jog past”
Unable to perceive ‘stereo sound’	Listening to the television	“I’m watching a programme that’s got multiple voices on the television occasionally I turn to my wife and say, “What did you say?” She hasn’t spoken but somebody on the TV has said something and I think it’s my wife joining in”
		“if you’re watching the television and you hear a voice off the screen, a female voice in my case, and I turn round to ask my voice what she said, because I didn’t hear what she said because I was concentrating on that, “I didn’t say anything.””
		“sit listening to a television programme and then there’s a dialogue going on, get somebody to speak in the room”
		“music because I can’t hear it like I used to”

	Listing to stereo music	"stereo music"
		"listening to music"
		"You could give them a piece of stereo music on headphones and have one headphone not working"
		"Or, going back to the spoken word, just have a piece of radio recording where they've got it in stereo and where the interviewer is in one speaker and the respondent is in another, with just a small element of transfer between them"
		"Music is actually produced so that some sound comes in one side and some comes in on the other. And if you can only hear in one ear that's really annoying because you only hear half of it."
		"it can be frustrating to hear something like Bohemian Rhapsody where it's one side, the other side, one side, the other side. You're not, you only get one side"
		"Kraftwerk. There's one particular track. It's called Autobahn and it's going from left to right in the speakers. Of this car racing down the road and to music as well but it's that could be a good one in terms of the sound is going to drop in left and right."
Miscellaneous	Softly spoken individuals	"another problem, [Name], you're very quietly spoken and that is a killer"
		"can't cope with his quiet speech"
	Not hearing a stranger	"you're sitting on a bus and the person next to you says, "Excuse me" quite quietly and you don't hear them, then they get louder and louder, then they think you're just being really rude"

		"Participant 1: When they want to get off, yes, I've had that. Participant 3: Getting angry with you because they think you're being rude but you just can't hear them"
		"there was somebody there saying excuse me three or four times and my husband said somebody wants to get by. I hadn't got a clue that she had"
	Generally outdoor noise	"there are times when I think am I going deaf? Because someone's in front of me and speaking to me. Especially if we are out of doors"
		"Just a little bit of wind can mess everything up"
		"Outside is difficult"
	Swimming pools	"the sound of the swimming pool. Quite echoey"
		"And the water and people shouting"
		"I hate swimming baths because of the echoey sound"

Table 32: Main themes and sub-themes identified within both focus groups with accompanying quotations.

7.1.3.8 Ranking exercise

The main themes and sub-themes identified from the focus groups can be found ranked in table 34 . Participants' individual rankings and the average ranking that the theme was allocated are included in the table, a category for 'other' was originally included in the ranking exercise, however this option was removed from the final ranking exercise as only four participants suggested items under this category and the situations detailed by participants had been covered within the discussions of the main themes in this research (difficulty listening to the television, or difficulty hearing passengers when driving). Impact was ranked on a seven-point scale with one being the most impactful and important to simulate and seven being the least impactful and important to simulate.

Main theme	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	Average Overall Ranking
Understanding what other people are saying in the presence of general background noise	1	2	3	1	2	1	1	1	3	6	2.10
Identifying the direction from which sounds are coming from, with or without noise	3	1	2	4	1	4	2	3	2	5	2.70
Listening to one person in a large group when more than one person is talking at any time	2	x	x	2	4	x	1	2	3	6	2.86
Hearing sounds located on the side of the impaired ear	4	5	1	3	5	2	3	4	1	4	3.20
Hearing and identifying unexpected sounds	6	4	6	5	3	3	4	6	4	1	4.20
Unable to perceive 'stereo sound'	5	2	7	7	6	5	7	5	5	x	5.44

Table 33: Main themes listed in rank order according to order of most important (lowest average score) to simulate to least important (highest ranking score).

7.1.4 Discussion

The main aim of this qualitative study was to identify specific situations that could be used to develop authentic, representative descriptions of the difficulties experienced by those living with SSD. Focus groups with individuals with SSD yielded multiple suggestions of possible scenarios to simulate for each of the four main themes (see table 33), with a minimum of two ASH options for each theme. Many of the findings of this research are in line with the findings of other studies that have investigated the main consequences of hearing loss. In particular, previous studies have found that adults with hearing loss experience difficulties with detecting and locating sounds, speech understanding in the presence of background noise, speech understanding during group conversations, listening to music and television, and communicating with unfamiliar people (Vas et al., 2017, Danermark et al., 2013).

Participants suggested a large number of possible scenarios for both the ‘identifying the direction from which sounds are coming from with or without noise’ and ‘hearing and identifying unexpected sounds’ themes. There are several instances of overlap with some suggested situations appearing within numerous themes, for example; ‘someone calling out your name’ is a suggested situation for both identifying the direction of sounds with and without noise and hearing sounds on the impaired side; a meeting scenario in a crowded room is a suggested situation for both identifying direction of sounds with and without noise and listening to one person in a large group; a conversation with other speakers present is a suggested situation for both understanding what other people are saying in the presence of general background noise and listening to one person in a large group; ‘crossing a road’ is a suggested situation for both identifying direction of sounds with and without noise, hearing sounds on impaired side, and shares aspects with the unexpected traffic noise scenario and bicycles and motorbikes passing, which were suggested situations to highlight difficulties hearing and identifying unexpected sound. There were several suggested situations which could be appropriate to simulate the effects of more than one theme, (e.g. dining in a restaurant scenario could be utilised to simulate

“Understanding what other people are saying in the presence of general background noise” or “Listening to one person in a large group when more than one person is talking at any time”) might be similar, but the difficulty within that situation (e.g. speech perception, sound location) might be different. The suggested ASH situations described by participants were consistent across both focus groups, despite participants being different ages, jobs and having a diagnosis of SSD for varying time frames and with varying origins, and the overlap of themes across different scenarios shows the consistent and far-reaching effects of SSD in daily life.

When ranking the themes and their impact and importance to simulate ‘understanding what other people are saying in the presence of general background noise’, ‘identifying the direction from which sounds are coming from with or without noise’, and ‘listening to one person in a large group when more than one person is talking at any time’ all scored under 3, indicating a high impact and importance. However, this rudimental ranking exercise was indicative of a conceptual impact, and should any new ASH be developed they would benefit from being evaluated by SSD participants and then again ranked more rigorously once a more developed schematic is produced.

7.1.5 Limitations

One of the limitations of this study was that convenience sampling was used, which is one of the least rigorous qualitative sampling techniques (Patton, 1990.). Unfortunately, convenience sampling was necessary due to the limited resources of this research. Specifically, time and money restrictions meant that it was not possible to use multiple recruitment avenues in addition to the NHBRU database, such as by placing advertisements in newspapers or by obtaining permission to train audiologists in NHS clinics to recruit participants for the study. Another potential limitation of this study was that the content analysis was deductive, as the themes were derived from the findings of Lucas et al. (*in prep*). Deductive analysis is a “top-down” approach that is rooted in a pre-existing framework or the researcher’s analytical interests. In contrast, inductive analysis is a “bottom-up” or data-driven process (Braun and Clarke 2006). It can be argued that deductive analysis increases the risk of overlooking important aspects of the difficulties experienced by individuals with SSD that do not match the pre-specified themes. However, even inductive analysis can be influenced by the researcher’s preconceptions and their knowledge of the relevant literature and theories (Malterud, 2001; Braun and Clarke, 2006). The deductive approach was selected for this study because it is considered suitable for fulfilling a specific research aim, whereas the inductive approach is better suited to exploring the data to develop new research questions (Braun and Clarke 2006). The semi-structured style used in the focus groups gave participants the opportunity to raise topics that were not directly relevant to the pre-existing themes but that were potentially important for this research.

7.1.6 Implications

This study produced a series of situations that are representative of the difficulties experienced by individuals with SSD in their everyday lives. The primary benefit of the qualitative approach adopted in this study is that these situations are likely to have strong content validity and comprehensibility, as they were produced in partnership with adults with SSD.

The findings of this study have implications for clinical practice. In particular, hearing healthcare professionals could use the authentic scenarios generated by individuals with SSD during history-taking in order to assess the difficulties experienced by their patients with SSD in a comprehensible, thorough, and patient-centred manner. In turn, this approach could help hearing healthcare professionals to identify the most appropriate intervention(s) to address the specific difficulties experienced by each of these patients.

The findings of this study also have implications for research. The overlap of scenarios generated by this study could prove to be advantageous, as it suggests that only a small number of new ASH are required to produce a more complete view of listening with SSD. Therefore, the extended version of the ASH tool would not be overly burdensome or onerous to develop.

Two of the suggested scenarios share similarities with the two ASH already developed as a part of this research. Firstly, dining in a restaurant and a meeting in a crowded room mirrors the current SiS+B ASH (see section 3.2.2), in which there is a target talker in front of the listener who speaks to the listener, while two further speakers have a back and forth exchange to the listener’s left, the unimpaired side, and general background noise (i.e. other conversations, banging noises, tableware moving) is present. Secondly a conversation with other speakers present shares similarities with the current SiS ASH (see section 3.2.2), in which there is a target talker in front of the listener who speaks to the listener while two further speakers have a back and forth exchange to the listener’s left, their unimpaired side. The new scenarios discussed within the focus groups would be relatively straight-forward to initially draft as the rich data provided by the focus group participants could be used as the basis of new simulations. Descriptions and schematics of the situations could then be re-presented to SSD patients for further development and clarification. Any additional simulations could be made using computer programming, as previous ASH have been built or through the use of recording devices in real world environments to further improve ASH and their ability to describe a hearing-related health state and its impact on individuals QoL.

7.1.7 Conclusions

This study aimed to identify situations that are representative of the difficulties experienced by individuals living with SSD in order to inform the further development of the ASH. A qualitative approach, which entailed consulting individuals with SSD, was used to ensure that these situations would be authentic and comprehensible. The findings provided rich and novel insights into the main challenges encountered by individuals with SSD, which included difficulties with understanding others in the presence of background noise, identifying the direction from which sounds are coming, listening to one person in a large group conversation, and hearing unexpected sounds or sounds located on the side of the impaired ear. These findings can inform clinical practice, particularly history-taking, as well as future research.

8 Discussion

This PhD project was the first development and validation of ASH to describe a hearing-related health state, providing empirical evidence of the validity and reliability of ASH as a descriptive tool, and establishes whether ASH could be appropriate for use in economic evaluation research.

The specific objectives of this research were:

(v) To assess the validity of using ASH to describe a hearing-related health and to value the quality of life of health states related to hearing loss

(vi) To assess the reliability of quality of life valuations obtained using ASH

(vii) To explore how best to use and present ASH when collecting health state preference data for health states related to hearing loss

(viii) To explore what situations patients think should be simulated when using ASH to describe health states related to hearing loss

8.1 Key findings

8.1.1 The validity of using ASH as a tool to describe health states related to hearing loss

As this research project was in its early stages two rudimentary ASH of SiS and SiS+B were built, based on vignettes developed within the Cochlear Implantation team which examined the subjective psychological and social effects of SSD in adults. a text-based description of living with SSD which was fully validated by both SSD patients and audiologists was also developed within the scope of this work (see appendices 9.1). In terms of simulating the effects of a hearing condition initially questions centred around the suitability of ASH as a health description tool, would it even be possible to convey aspects of listening difficulties that individuals living with SSD experience? Could ASH convey these difficulties as a stand-alone method, or would the ASH need to be presented with a complementary piece of text?

Pilot testing showed that a basic, rigid simulation of SSD was a feasible and that such simulations could communicate different levels of difficulty that can be experienced by someone living with hearing loss, to participants with healthy hearing. Utility values collected using ASH to describe an SSD hearing-related health state were less than the utility value given for living in perfect health and an effect of listening condition was indicated with the average SiS+B health utility values lower than those for SiS, this is in line with the literature which reports that adults with hearing loss experience difficulties with speech understanding in the presence of background noise, and speech understanding during group conversations (Arlinger, 2003, Vas et al., 2017, Danermark et al., 2013). Finally, a difference across modalities; between audio, ASH and text and text only, was not obvious until the ASH had been redeveloped with a new sentence corpus and less predictability (i.e. the introduction of jitter to remove predictability of talker timings from the ASH). The ASH became more robust after version 1 ASH were altered and improved to sound more naturalistic by using the BKB sentences to replace the more rigid CRM sentence corpus and reducing some level of predictability from the ASH. Utility valuations collected within experiments 4, 5 and 6 which used ASH as a descriptor alongside text were similar to those collected from SSD patients (Summerfield et al 2002, 2010). The fact that utilisation of ASH within the health-state valuation process has consistently resulted in successful collection of utility values highlights ASH are a valid health state description tool. Also utility values collected throughout this body of work were aligned with valuations given by individuals with SSD.

8.1.2 The reliability of ASH as a tool to describe health states related to hearing loss.

In experiment 1 presenting ASH alone with no supplementary text resulted in the health-state utility values collected lacking reliability and susceptible to influence from other sources of information (i.e. any text-based descriptions presented in other stages of experiments). Within experiment 1 when comparing the first evaluation of the SSD health state collected using ASH alone, when the participant's decisions have not yet been influenced by any information in the text, and the second evaluation of the same health state, there was evidence from the poor ICC scores that participants were changing their minds about the QoL of somebody living with the SSD described in the ASH. All ICC values and their confidence intervals for experiment fell below the level of good reliability identified by Koo and Li (2016). This failure to achieve sufficient reliability suggests that the participants could have been influenced by the text descriptions presented within session 1 of this experiment.

However, the low levels of reliability observed could not entirely be attributed to the influence of the text-based descriptions on the utility valuations participants gave, it could also have been attributed to the lack of familiarity the participants had with the valuation process in particular the TTO and VAS tasks, lack of familiarity with the ASH themselves and the ASH not accurately simulating the listening difficulties associated with SSD, a lack of familiarity with the hearing condition described and not receiving more information due to a lack of content validity within the ASH. It is also worth noting that all other presentation methods also showed varying levels of reliability. To improve reliability ASH were redeveloped to be more realistic, more examples were given to the participants about how they should interact with the TTO and VAS tasks, and the TTO and VAS sliders were made more sensitive, with TTO intervals reduced from six months to a month (see section 3.5)

8.1.3 How best to use and present ASH within an economic evaluation

Throughout this body of work how exactly to present ASH to participants has been explored. From experiment 1 it became apparent that the health state valuation process is best presented in a single session. There were low levels of reliability for the health utility values collected between experiment 1 testing sessions for all presentation methods. This calls into question the usefulness of collecting valuations in two separate sessions, especially as collecting economic evaluations over two separate sessions is not very practical. The scheduling and time demands of two testing sessions was excessive, this could reduce feasibility of using ASH widely in future to collect utility values. Separate testing sessions were not used beyond experiment 1.

Throughout experiments 4 and 5 the redeveloped version 3 ASH were presented repeatedly to investigate the reliability of ASH over time, it was concluded that the utility values collected using ASH are reliable from the first presentation, meaning that multiple presentations of the same health state is not necessary. This is beneficial as it means that a valuation experiment using ASH is not incredibly long and arduous for the participants to complete.

The presentation of ASH alongside an active listening task and with a comparator simulation of perfect health was also investigated in chapter 5 (5.4, 5.5). These 2 experiments showed no significant difference between utility values when presented with a task nor a perfect health comparison simulation, however, adding such measures, especially if the valuation process is carried out under less controlled conditions or remotely as was the case in experiment 8, could encourage participants to fully engage with the valuation process. This extra level of control could give a level of reassurance to the researcher that the participant rating the health state is in fact interacting with the ASH and has a comparison between the hearing-related health state and a normal-hearing health state should they need clarification. Finally, presenting an ASH of a hearing-related health state and an intervention-related health state, i.e. a CI was explored briefly. Experiment 7 and 8 indicated that it is possible to use ASH to describe treatment effects. Unfortunately due to time constraints it was not possible to delve into the intricacies of how other treatment options could be simulated, such as a hearing aid, future research could explore how best to simulate a variety of treatment options available for individuals with hearing loss.

8.1.4 The development of additional ASH to describe SSD

Within this research two ASH were developed and then improved upon, after the initial development of two main exemplar simulations, no further simulations were made. Qualitative work detailed in Chapter 7 provided ideas and details for other simulations that could be beneficial to include in utility task using ASH. Multiple new simulation ideas were detailed by patients during the focus groups (see chapter 7). It may also be possible to use real world recordings of staged situations, much like Kayser et al (2009) did when developing the cafeteria simulation used within this research. The real-world recordings could be taken using an in-ear microphone, in real world settlings and then digitally manipulated. It would take time to build and revalidate these new ASH scenarios by going back to SSD participants and possibly audiologists with details and schematics, but it would ensure high content validity of any additional ASH.

8.1.5 Remote presentation of ASH using a web-based platform

Presenting ASH and the valuation process online was investigated as a way to reduce the burden on research resources when collecting valuation data from large groups of people. The collection of utility values was successful with valuations collected remotely; these valuations were also in line with valuations collected in person. Presenting valuation tasks online can mean a shorter time frame for the collection of responses and be time and cost saving, the valuations for experiment 7 were collected within 22 days. Utilising an online task via the world wide web also gives researchers access to a larger, potentially more diverse participant population and the prospect of collecting larger amounts of data (Lefever et al 2007).

A researcher presenting a utility valuation task over the internet must remain aware that if participants have questions, concerns, or misunderstand with the task that the researcher is not immediately available to guide participants as they would in a face-to-face setting. When collecting data remotely the researcher has little control of the environment a participant is in or the equipment a participant is using, however with forethought it is possible to use supplementary tasks to ensure a basic level of adherence to study requirements, for example requests to use headphones within the study information section of the website and the inclusion of a headphone test within experiments presented online to prompt participants to use headphones. The use of listening tasks within experiments to try to encourage participants to engage with the ASH and encourage attention is directed to the valuation tasks.

Another factor to be considered when deploying online methods is participant attrition. The dropout rates that occurred during experiment 8 are something that must be taken into consideration when administering online research. For experiment 8 the dropout rate in terms of how many participants started the experiment and how many finished, was acceptable as although 21 did not complete the study 30 did. Where participants dropped out was not related to a specific point or section which indicates participants were not dropping out due to one aspect of the study that they could not complete. As a way to alleviate this limitation, future researchers may wish to recruit more than their sample size calculation dictates to account for drop out and may wish to investigate in more detail the factors affecting drop out within the valuation process, is it the length of the task itself, is it a lack of understanding of the TTO or VAS task, etc.

8.2 Strengths and limitations of the research

8.2.1 Strengths and limitations of ASH

The development of a validated and reliable set of ASH hinge upon the existence of a validated and reliable vignette of the hearing condition of interest. If a detailed and extensive, validated text description of the hearing-related health condition of interest is not yet available to sign post ASH development, i.e. situations which cause listening difficulty then ASH of that health condition would be difficult to develop. Further validation work would be needed to ensure the simulations displayed adequate content validity, which would be a time-consuming process to undertake.

The ASH developed within this research required multiple small changes and revisions to ensure optimum content validity. It may also be beneficial to delve into the optimum length for a simulation, this research assumed the optimum length of the ASH for practical reasons, ASH were kept relatively short by a lack of talkers within BKB the sentence corpus and a lack of resources and time needed to recruit and investigate optimum ASH length. However, once ASH have been designed and developed, they are a versatile tool which can be presented remotely, used to present interventions such as CI and that requires little resources to present to participants.

8.2.2 Strengths and limitations of methodologies

It is important to note that although this research was investigating how ASH could be utilised within the economic evaluation of hearing-related health states this was not the sole focus. The research also aimed to develop ASH and a large amount of time was given to this task. Initially the use of valuation methodology, namely TTO and VAS tasks in conjunction was guided by the literature. Both TTO and VAS methodologies have been used within economic evaluation research, both TTO and VAS were used in this research in an attempt to increase the rigor of the experiment and ensure some form of data could be collected as ASH were novel and it was unknown as to whether it would be possible to successfully collect preference based utility values using ASH and TTO. However as VAS is not a preference-based method of collecting health utility values, to improve the robustness of any future research it may be prudent to collect valuations using only TTO which is a preference-based method of collecting health utility values.

Using the HUI descriptions of perfect, deaf, and worst health states was useful for this research. The HUI coverage is restricted to physical and emotional aspects of health and items cover vison, hearing, speech, ambulation, dexterity, emotion, cognition, and pain. It is important to note that there is a limitation of applying the HUI to economic evaluations in the context of the UK as the utility values it provides are based on the preferences of a sample of the Canadian public. However the HUI3 is viewed as more sensitive to hearing-related health states (Longworth et al., 2014).

Along with some limitations present with the types of methodologies used the representativeness of participants who took part in the research should be noted. Many of the participants were students, any future research should try to actively recruit non-students in order to assure a representative sample. This was not possible within this research due to the utilisation of opportunity sampling and time and funding restraints.

8.3 Recommendations for future research in this area.

8.3.1 Future directions for ASH

The development of other ASH representative of other listening situations would be beneficial in the future as the use of multiple ASH to demonstrate many facets of momentarily hearing with a hearing loss alongside a vignette or health valuation questionnaire could improve the understanding of someone valuing a hearing-related health state without relying on the participant’s conceptualisation abilities. The use of a larger compliment of ASH to describe a hearing-related health state would need to be investigated by future research and the method validated as the two ASH developed during this research were.

It could be possible and advantageous for further research to simulate not only SSD but also other hearing loss types, such as mild-to-moderate hearing loss or tinnitus, or to use currently available computer simulation software and methods such as MatLab to apply the simulation methodology to visual conditions, such as tunnel vision, or cataracts etc.

There are also several different avenues that ASH could be taken down besides use within the description of a health state within the valuation of QoL. With the emergence of new technologies such as VR and given their current rate of improvements, it is not infeasible to suggest that virtual reality may soon be used to simulate many more conditions such as physical disabilities, all types of hearing conditions, visual conditions, vestibular conditions etc.

There may also be potential for ASH to be used for educational purposes, such as within the school system, or within workplaces for employers, to show people who work with and interact with individuals with a hearing loss how that hearing loss affects them. The same sort of service could also be offered to the family members of individuals with a recent diagnoses of SSD so that family members can better understand the impact the hearing loss has on their lives.

8.3.2 Using simulations within the economic evaluation process

The ASH developed within the research are still relatively rudimentary, nonetheless with extra development and validation they could become a vital tool for describing a hearing-related health state. This extra work is not infeasible either, researchers spend many hours and use countless resources developing and validating questionnaires and tools such as the HUI (which is currently in its third iteration since initial conception) and the EuroQoL group has developed multiple versions of their questionnaire; EQ-5D with 3 levels of severity for each of the 5 dimensions: EQ-5D-3L, EQ-5D with 5 levels of severity for each of the 5 dimensions: EQ-5D-5L, EQ-5D for use in children: EQ-5D-Y.

ASH or other types of simulations could be the next step in the evolution of health state description and valuations. This could be as a supplementary resource for vignettes describing health states or as an addition to health valuation questionnaire much the way ‘bolt-ons’ to the EQ-5D have been suggested as a way of tailoring the tool towards increased sensitivity for issues such as vision and hearing impairments.

Before ASH can be successfully presented with a health valuation questionnaire or vignette it would be necessary to validate the method using a larger sample size calculated using a power calculation with a higher power than the 80% power used within these experiments. The sample sizes used in this research were suitable and feasible for development and validation work but a full economic evaluation of SSD with a text-based vignette and additional ASH would require more participants, time resources than were readily available during this research.

8.4 Conclusion

ASH has a huge potential as a descriptive tool both within the field of hearing and other subjective conditions, as does the simulation of health states using current and developing technologies such as virtual reality. ASH could provide benefit outside the realm of economic evaluation, with the potential for ASH to be utilised as an educational instrument to describe living with SSD and other hearing conditions when educating employers, education staff, and family members of patients about the effects of living with SSD. Although ASH take a significant amount of time, effort and resources to develop they could revolutionise health valuation process, especially in light of the advancements in technologies, such as virtual reality. Describing the impact of a health condition using simulations could well be the key to accurate QoL valuations that are in line with how patients feel and self-report, without relying on conceptualisation abilities of the rating public.

9 Appendices

9.1 Text-based description of Single-Sided Deafness

You have a profound hearing loss in your left ear. This means that you cannot hear anything using your left ear. You have little or no hearing loss (‘normal’ hearing) in your right ear. You only hear sounds through your right ear. People do not recognise that you have a hearing loss because you don’t wear a hearing aid. Friends and family often forget about your hearing loss and will talk to you on the side of your deaf ear. You often have to explain to strangers about the consequences of your hearing loss and remind them to position themselves on your hearing side or in front of you which helps you to hear. Sometimes you find that people make little effort to talk to you because they do not understand your hearing difficulties.

You cannot identify where sounds are coming from, particularly unexpected sounds, which you find frustrating. When crossing the road you find it difficult to identify the direction that traffic is travelling and how far it is away from you. You are sometimes surprised that traffic is closer than you first thought. You rely on visual clues to identify where the source of a sound is located.

You have difficulty following conversations in a noisy room especially when there are several conversations taking place at once. In a quiet room, you have little or no difficulty hearing and understanding others as long as one person is speaking at any time. In larger groups of people, you find it difficult to follow conversations because you do not always know who has spoken. You sometimes worry that other people may find you rude or ignorant because you do not respond to their direct conversation or you miss what they have said. You rely on visual clues and lip reading to help you participate in conversations.

In noisy environments, you try to position yourself in a corner of a room so that the noise does not completely surround you. Regardless of where you position yourself, you sometimes struggle to follow conversations. It is difficult for you to participate in conversations in places with lots of background noise such as at a pub or café, so you sometimes make excuses to avoid these situations. However, you enjoy socialising and don't want to miss out, so often agree to go anyway. You often struggle to join in conversations whilst you are there making you feel frustrated and isolated. You find it almost impossible to hear in places where there is a very high level of background noise and echoes such as at train stations or airports. In these situations, you rely on other people to ensure that you do not miss important information.

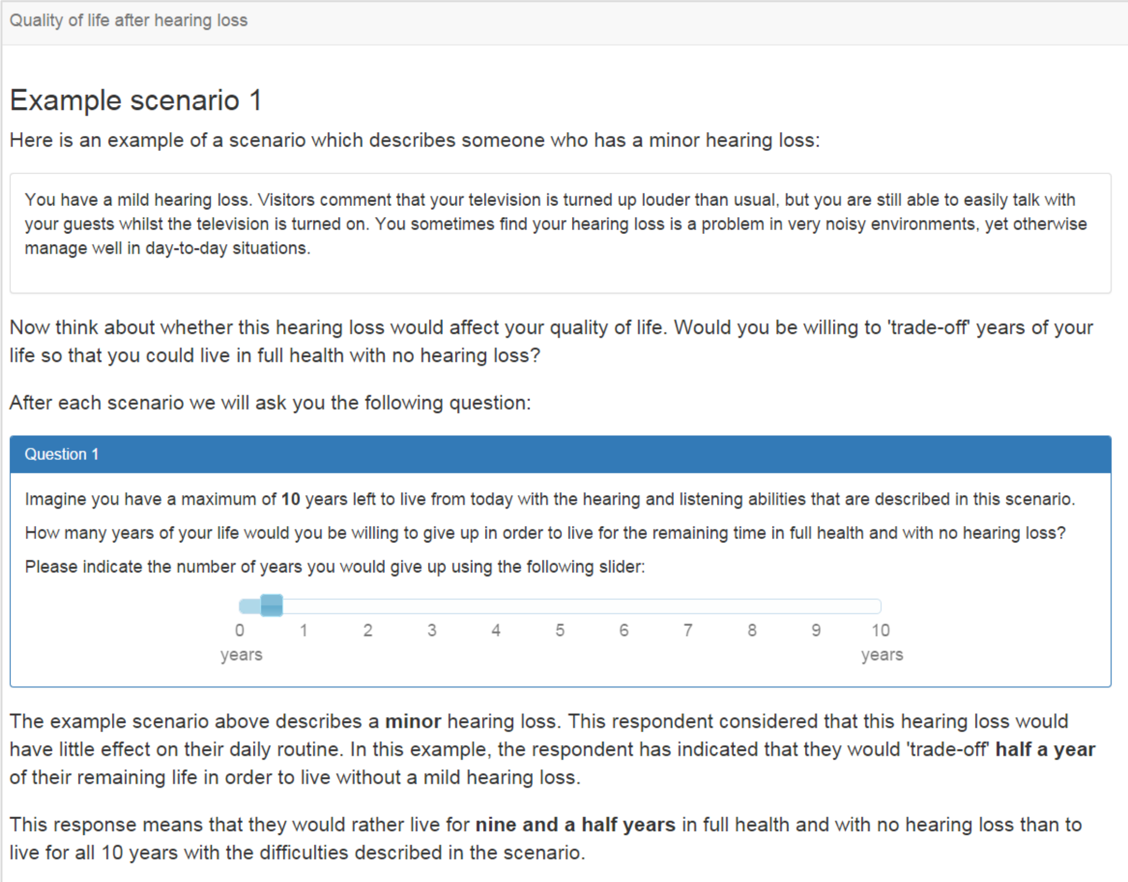
The overall effort of listening on a day-to-day basis can be tiring for you, especially when you have to listen in more challenging situations such as in background noise. You find that turning your head to hear people can be very uncomfortable. You are often exhausted at the end of the day and get tired more easily than before your hearing loss. People can mistake this to mean that you are being antisocial, moody, or rude. The reaction of other people upsets and frustrates you.

Occasionally, you find that other people don't understand the consequences of your hearing loss. When you miss part of a conversation, particularly when it is with strangers or work colleagues, you frequently just nod in agreement instead of asking people to repeat themselves several times because you find this can be embarrassing.

Sometimes people can be rude to you when you haven't heard what they have said which leaves you feeling annoyed. Your close friends and family have a better understanding of your hearing loss and will make adjustments to ensure you can join in conversations such as repeating what has been said, talking more loudly, shouting your name to get your attention, or positioning themselves on your good side.

You hear ringing, buzzing, or other kinds of sounds in your left ear ('tinnitus'). Your tinnitus is most noticeable in quieter environments. Your tinnitus can make it difficult to concentrate on what you are doing. You may take precautions to protect your right ear from hearing damage, such as wearing an earplug in noisy situations or avoiding places where you know there will be loud music. You feel very upset and worried if you think about what would happen if you lost the hearing in your other ear.

9.2 Screenshots of experiment 1 interface



Quality of life after hearing loss

Example scenario 1

Here is an example of a scenario which describes someone who has a minor hearing loss:

You have a mild hearing loss. Visitors comment that your television is turned up louder than usual, but you are still able to easily talk with your guests whilst the television is turned on. You sometimes find your hearing loss is a problem in very noisy environments, yet otherwise manage well in day-to-day situations.

Now think about whether this hearing loss would affect your quality of life. Would you be willing to 'trade-off' years of your life so that you could live in full health with no hearing loss?

After each scenario we will ask you the following question:

Question 1

Imagine you have a maximum of **10** years left to live from today with the hearing and listening abilities that are described in this scenario.

How many years of your life would you be willing to give up in order to live for the remaining time in full health and with no hearing loss?

Please indicate the number of years you would give up using the following slider:

The example scenario above describes a **minor** hearing loss. This respondent considered that this hearing loss would have little effect on their daily routine. In this example, the respondent has indicated that they would 'trade-off' **half a year** of their remaining life in order to live without a mild hearing loss.

This response means that they would rather live for **nine and a half years** in full health and with no hearing loss than to live for all 10 years with the difficulties described in the scenario.

Example scenario 2

Here is an example of a different health state. This person has experienced a stroke:

You have had a stroke. You have lost the ability to control your hands and to speak clearly. You are regularly forgetting to do things. As a result, you are struggling to cope with day-to-day living. Your condition is stable, and will neither get better, nor get worse.

Question 1

Imagine you have a maximum of **10** years left to live from today with the health described in this scenario.

How many years of your life would you be willing to give up in order to live for the remaining time in full health?

Please indicate the number of years you would give up using the following slider:

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

In our experience, most people consider the consequences of a stroke to be a **major** health problem. In the example shown here, the respondent has decided they would be willing to 'trade-off' **seven years** of their life to live in full health.

This response means that they would rather live for **three years** in full health than to live for all 10 years with the difficulties described in the scenario.

Next

27%

This study is being conducted by the NIHR Nottingham Hearing Biomedical Research Unit. Click here for more information.

To help you understand what living with hearing loss is like, we have prepared two demonstrations for you to listen to. The first demonstration will give you an idea of what it is like to listen to a conversation in a quiet environment with good hearing; i.e. with no hearing loss. Please listen to this demonstration carefully and repeat it as many times as you like:

00:00 -00:11

Demonstration of listening without hearing loss

The second demonstration will help you imagine what it is like to listen in this same environment but with a hearing loss. Please listen to the following demonstration carefully and imagine that it demonstrates your **own** ability to hear today:

00:00 -00:11

Demonstration of listening with hearing loss

Question 1

Imagine you have a maximum of **10** years left to live from today with the hearing and listening abilities that have been demonstrated using the audio examples above.

How many years of your life would you be willing to give up in order to live for the remaining time in full health and with no hearing loss?

Please indicate the number of years you would give up using the following slider:

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

9.3 Experiment 3 screenshots

Scenario 4

Please read the following text imagining that it describes your **own** health today.

You lost all hearing in your right ear several years ago. You have no hearing loss in your left ear.

You cannot identify where sounds are coming from. All sounds appear to come from your left side.

You try to avoid places with lots of background noise such as pubs or cafés. Having a conversation in these places can be difficult, frustrating and isolating.

You are often exhausted at the end of the day, especially after listening in background noise.

To help you understand what living with this type of hearing loss is like, we have prepared two demonstrations for you to listen to. The first demonstration will give you an idea of what it is like to listen to a conversation in a noisy environment with good hearing; i.e. with no hearing loss. Please listen to this demonstration carefully and repeat it as many times as you like:

00:00 -00:12

Demonstration of listening without hearing loss

The second demonstration will help you imagine what it is like to listen in this same environment but with the hearing loss described above. Please listen to the following demonstration carefully and imagine that it demonstrates your **own** ability to hear today:

00:00 -00:12

Demonstration of listening with hearing loss

Question 1

Imagine you have a maximum of **10** years left to live from today with the hearing and listening abilities that have been demonstrated using the audio examples above.

How many years of your life would you be willing to give up in order to live for the remaining time in full health and with no hearing loss?

Please indicate the number of years you would give up using the following slider:

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

Saved

Question 2

To help us to gain a better understanding of how you rate your quality of life in these scenarios, we have drawn a scale.

The **best** quality of life you can imagine, with no hearing loss, is valued as **100**. The **worst** quality of life you can imagine is valued at **0**.

We would like you to choose a value on this scale to show how good or bad you believe your quality of life is in the scenario that was demonstrated using the audio examples.

0 Worst 10 20 30 40 50 60 70 80 90 100 Best

45%

This study is being conducted by the NIHR Nottingham Hearing Biomedical Research Unit. Click here for more information.

9.4 Vignette sample used for experiment 3, 4 and 5

Please read the following text imagining that it describes your **own** health today.

You lost all hearing in your right ear several years ago. You have no hearing loss in your left ear.

You cannot identify where sounds are coming from. All sounds appear to come from your left side.

You try to avoid places with lots of background noise such as pubs or cafés. Having a conversation in these places can be difficult, frustrating and isolating.

You are often exhausted at the end of the day, especially after listening in background noise.

9.5 Full text vignette used for experiment 6

Scenario 16

We would now like you to **re-read** the same text description again. However, this time we have also provided an **audio demonstration** to help you imagine what it is like to have this form of hearing loss.

Remember, we would like you to imagine that the text and the audio demonstration describe your **own** health today. Please re-read the text below, listen to the audio demonstration, and then re-answer the questions that follow.

You lost all hearing in your left ear several years ago. Your right ear has no hearing loss. You feel very upset if you think about losing the hearing in that ear.

You cannot identify where sounds are coming from. All sounds appear to come from your right side. When crossing a road, you cannot identify the direction that traffic is travelling in and how far away it is without looking around.

You can have a conversation in a small group as long as only one person talks at a time. You find it harder to follow the conversation if you cannot see everyone.

You try to avoid places with lots of background noise such as pubs or cafés. Having a conversation in these places can be difficult, frustrating, and isolating.

To hear in noisy places, you need to put people on your right side or sit in a corner so that noise does not completely surround you. You find it difficult to understand people if you cannot see their lips.

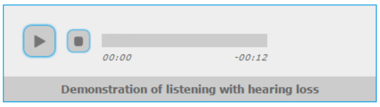
You often miss parts of conversations and worry that people may find you rude. You feel too embarrassed to ask people to repeat themselves several times.

You rely on other people to ensure you don't miss important information in train stations and airports because the noises and echoes make hearing almost impossible.

You are often exhausted at the end of the day, especially after listening in background noise.

Here is an audio example to help you imagine what it is like to listen with the hearing loss described above. In this demonstration, you are sitting at a table with three other people who are speaking.

Please listen to the following demonstration carefully. The **person sitting in front of you** will say **three** sentences. You task is to try to hear and remember the **third** sentence that they say.



9.6 CI vignette sample used for experiments 7 and 8

Please read the following text imagining that it describes your **own** health today.

You lost all hearing in your right ear several years ago. You have no hearing loss in your left ear. You had surgery to receive a cochlear implant (CI) which restored some hearing in your right ear.

You can tell whether a sound is on your left or right but you have difficulty identifying its exact location.

You will agree to meet in places with lots of background noise such as pubs or cafés. However, having a conversation in these places can sometimes be difficult and frustrating.

You find it easier to listen when using your CI. However, you still get tired at the end of the day, especially after listening in background noise.

References

1969. IEEE Recommnded Pratice for Speech Quality Measurements. *IEEE Transactions on Audio and Electroacoustics,* 17**,** 225-246.

AKEROYD, M. A. 2006. The psychoacoustics of binaural hearing: La psicoacústica de la audición binaural. *International journal of audiology,* 45**,** 25-33.

ALTMAN, D. G. & BLAND, J. M. 1983. Measurement in medicine: the analysis of method comparison studies. *Journal of the Royal Statistical Society: Series D (The Statistician),* 32**,** 307-317.

ANDERSON, A. K. 2005. Affective Influences on the Attentional Dynamics Supporting Awareness. *Journal of Experimental Psychology: General,* 134**,** 258-281.

ARLINGER, S. 2003. Negative consequences of uncorrected hearing loss--a review. *International journal of audiology,* 42 Suppl 2**,** 2S17-20.

ARNDT, S., ASCHENDORFF, A., LASZIG, R., BECK, R., SCHILD, C., KROEGER, S., IHORST, G. & WESARG, T. 2011. Comparison of pseudobinaural hearing to real binaural hearing rehabilitation after cochlear implantation in patients with unilateral deafness and tinnitus. *Otology & neurotology,* 32**,** 39-47.

ARNESEN, T. & TROMMALD, M. 2005. Are QALYs based on time trade-off comparable?--A systematic review of TTO methodologies. *Health Economics,* 14**,** 39-53.

ASCH, D. A., JEDRZIEWSKI, M. K. & CHRISTAKIS, N. A. 1997. Response rates to mail surveys published in medical journals. *Journal of clinical epidemiology,* 50**,** 1129-1136.

ATKINSON, G. & NEVILL, A. M. 1998. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports medicine,* 26**,** 217-238.

ATTEMA, A. E., EDELAAR-PEETERS, Y., VERSTEEGH, M. M. & STOLK, E. A. 2013. Time trade-off: one methodology, different methods. *The European Journal of Health Economics,* 14**,** 53-64.

AUDIOLOGY, B. S. O. 2011. Recommended procedure: Pure-tone air-conduction and bone-conduction threshold audiometry with and without masking. *Reading: British Society of Audiology*.

BANSOD, A., SKOCZEN, S. & LENERT, L. A. 2003. IMPACT4: a framework for rapid, modular construction of web-based patient decision support systems and preference measurement tools. *American Medical Informatics Association Annual Symposium proceedings*, 782.

BARTLETT, J. & FROST, C. 2008. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. *Ultrasound in Obstetrics and Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology,* 31**,** 466-475.

BARUCH, Y. 1999. Response rate in academic studies-A comparative analysis. *Human relations,* 52**,** 421-438.

BARUCH, Y. & HOLTOM, B. C. 2008. Survey response rate levels and trends in organizational research. *Human Relations,* 61**,** 1139-1160.

BENCH, J., KOWAL, A. & BAMFORD, J. 1979. The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *British Journal of Audiolgy,* 13**,** 108-12.

BHATTACHARYA, J., HYDE, T. & TU, P. 2014. *Health economics*.

BLAND, M. 2015. *An introduction to medical statistics*, Oxford University Press (UK).

BOKARI, S., PREPAGERAN, N. & RAMAN, R. 2010. Visual analog scale in hearing loss. *Indian Journal of Otolaryngology and Head & Neck Surgery,* 62**,** 40-3.

BOLIA, R. S., NELSON, W. T., ERICSON, M. A. & SIMPSON, B. D. 2000. A speech corpus for multitalker communications research. *The Journal of the Acoustical Society of America,* 107**,** 1065-1066.

BORTON, S. A., MAUZE, E. & LIEU, J. E. 2010. Quality of life in children with unilateral hearing loss: a pilot study. *American Journal of Audiology,* 19**,** 61-72.

BRAUN, V. & CLARKE, V. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology,* 3**,** 77-101.

BRAZIER, J., RATCLIFFE, J., SALOMON, J. & TSUCHIYA, A. 2007. *Measuring and Valuing Health Benefits for Economic Evaluation,* Oxford, Oxford University Press.

BRAZIER, J. & ROWEN, D. 2011. NICEDSU TECHNICALSUPPORT DOCUMENT 11: ALTERNATIVESTO EQ-5D FOR GENERATING HEALTH STATE UTILITY VALUES.

BROD, M., TESLER, L. E. & CHRISTENSEN, T. L. 2009. Qualitative research and content validity: developing best practices based on science and experience. *Quality of life research,* 18**,** 1263.

BROWN, K. D. & BALKANY, T. J. 2007. Benefits of bilateral cochlear implantation: a review. *Current opinion in otolaryngology & head and neck surgery,* 15**,** 315-318.

BUXTON, M. J. 2006. Economic evaluation and decision making in the UK. *Pharmacoeconomics,* 24**,** 1133-1142.

CADMAN, D. & GOLDSMITH, C. 1986. Construction of social value or utility-based health indices: the usefulness of factorial experimental design plans. *Journal of chronic diseases,* 39**,** 643-651.

CHENG, A. K. & NIPARKO, J. K. 1999. Cost-utility of the cochlear implant in adults: A meta-analysis. *Archives of Otolaryngology–Head & Neck Surgery,* 125**,** 1214-1218.

CHENG, A. K., RUBIN, H. R., POWE, N. R., MELLON, N. K., FRANCIS, H. W. & NIPARKO, J. K. 2000. Cost-utility analysis of the cochlear implant in children. *Journal of the American Medical Association,* 284**,** 850-856.

CHIOSSOINE-KERDEL, J. A., BAGULEY, D. M., STODDART, R. L. & MOFFAT, D. A. 2000. An investigation of the audiologic handicap associated with unilateral sudden sensorineural hearing loss. *American Journal of Otology,* 21**,** 645-51.

CLARK, G. 2004. Cochlear implants. *Speech processing in the auditory system.* Springer.

CLARKE, A., GOLDSTEIN, M., MICHELSON, D., GARBER, A. & LENER, L. 1997. The effect of assessment method and respondent population on utilities elicited for Gaucher disease. *Quality of Life Research,* 6**,** 0-0.

CLAYTON, P. A. & MACKAY, D. P. 2017. Experienced Utility or Decision Utility for QALY Calculation? Both. *Public Health Ethics,* 11**,** 82-89.

DANERMARK, B., GRANBERG, S., KRAMER, S. E., SELB, M. & MOLLER, C. 2013. The creation of a comprehensive and a brief core set for hearing loss using the international classification of functioning, disability and health. *Am J Audiol,* 22**,** 323-8.

DAVIS, A. & RESEARCH, M. R. C. I. O. H. 1995. *Hearing in adults : the prevalence and distribution of hearing impairment and reported hearing disability in the MRC Institute of Hearing Research's National Study of Hearing,* London, Whurr Publishers.

DE VET, H. C., TERWEE, C. B., MOKKINK, L. B. & KNOL, D. L. 2011. *Measurement in medicine: a practical guide*, Cambridge University Press.

DEVON, H. A., BLOCK, M. E., MOYLE-WRIGHT, P., ERNST, D. M., HAYDEN, S. J., LAZZARA, D. J., SAVOY, S. M. & KOSTAS-POLSTON, E. 2007. A psychometric toolbox for testing validity and reliability. *Journal of Nursing scholarship,* 39**,** 155-164.

DOLAN, P., GUDEX, C., KIND, P. & WILLIAMS, A. 1996. Valuing health states: a comparison of methods. *Journal of Health Economics,* 15**,** 209-31.

DOUGLAS, S. A., YEUNG, P., DAUDIA, A., GATEHOUSE, S. & O'DONOGHUE, G. M. 2007. Spatial hearing disability after acoustic neuroma removal. *Laryngoscope,* 117**,** 1648-51.

DOYLE, S., LLOYD, A. & WALKER, M. 2008. Health state utility scores in advanced non-small cell lung cancer. *Lung Cancer,* 62**,** 374-80.

DRUMMOND, M., SCULPHER, M., TORRANCE, G., O'BRIEN, B. & STODDART, G. 2005. *Methods for the economic evaluation of health care programme.,* Oxford, Oxford University Press.

DUSCH, M. N., O'SULLIVAN, P. S. & ASCHER, N. L. 2014. Patient perceptions of female surgeons: how surgeon demeanor and type of surgery affect patient preference. *Journal of Surgical Research,* 187**,** 59-64.

ELBERLING, C. & WORSOE, K. 2006. *Fading Sounds: about hearing and hearing aids,* Denmark, The Oticon Foundation.

FAUL, F., ERDFELDER, E., BUCHNER, A. & LANG, A.-G. 2009. Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods,* 41**,** 1149-1160.

FAUL, F., ERDFELDER, E., LANG, A.-G. & BUCHNER, A. 2007. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods,* 39**,** 175-191.

FEENY, D., FURLONG, W., TORRANCE, G. W., GOLDSMITH, C. H., ZHU, Z., DEPAUW, S., DENTON, M. & BOYLE, M. 2002. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Medical care,* 40**,** 113-128.

FERGUSON, M., LEIGHTON, P., BRANDRETH, M. & WHARRAD, H. 2018. Development of a multimedia educational programme for first-time hearing aid users: A participatory design. *International Journal of Audiology,* 57**,** 600-609.

FLORENTINE, M., FASTL, H. & BUUS, S. R. 1988. Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking. *The Journal of the Acoustical Society of America,* 84**,** 195-203.

GERHARDS, S. A., EVERS, S. M., SABEL, P. W. & HUIBERS, M. J. 2011. Discrepancy in rating health-related quality of life of depression between patient and general population. *Quality of Life Research,* 20**,** 273-9.

GIOLAS, T. G. & WARK, D. J. 1967. Communication problems associated with unilateral hearing loss. *Journal of Speech and Hearing Disorders,* 32**,** 336-343.

GOLDSTEIN, M. K., CLARKE, A. E., MICHELSON, D., GARBER, A. M., BERGEN, M. R. & LENERT, L. A. 1994. Developing and testing a multimedia presentation of a health-state description. *Medical Decision Making,* 14**,** 336-44.

GOLDSTEIN, M. K., MICHELSON, D., CLARKE, A. E. & LENERT, L. A. A multimedia preference-assessment tool for functional outcomes. Proceedings of the Annual Symposium on Computer Application in Medical Care, 1993. American Medical Informatics Association, 844.

GRAZIANO, A. M. & RAULIN, M. L. 2014. *Research methods : a process of inquiry,* Harlow, Pearson.

GRIFFITHS, P. & MURRELLS, T. 2010. Reliability assessment and approaches to determining agreement between measurements: classic methods paper. *International journal of nursing studies,* 47**,** 937-938.

GUDEX, C. 1994. Time trade-off user manual: props and self-completion methods.

HEFFERNAN, E., COULSON, N. S., HENSHAW, H., BARRY, J. G. & FERGUSON, M. A. 2016. Understanding the psychosocial experiences of adults with mild-moderate hearing loss: An application of Leventhal's self-regulatory model. *International Journal of Audiology,* 55**,** S3-S12.

HORSMAN, J., FURLONG, W., FEENY, D. & TORRANCE, G. 2003. The Health Utilities Index (HUI®): concepts, measurement properties and applications. *Health and quality of life outcomes,* 1**,** 54.

HOSFORD-DUNN, H., ROESER, R. J. & VALENTE, M. 2011. *AUDIOLOGY Practice Management*, Thieme.

HOUSE, W. F. 1976. Cochlear implants. *Annals of Otology, Rhinology & Laryngology,* 85**,** 3-3.

HOWITT, D. & CRAMER, D. 2007. *Introduction to statistics in psychology.* .

HSIEH, H. F. & SHANNON, S. E. 2005. Three approaches to qualitative content analysis. *Qualitative Health Research,* 15**,** 1277-88.

JANSEN, S. J., STIGGELBOUT, A. M., WAKKER, P. P., NOOIJ, M. A., NOORDIJK, E. M. & KIEVIT, J. 2000. Unstable preferences: a shift in valuation or an effect of the elicitation procedure? *Medical Decision Making,* 20**,** 62-71.

JEUB, M., SCHAFER, M. & VARY, P. A binaural room impulse response database for the evaluation of dereverberation algorithms. 2009 16th International Conference on Digital Signal Processing, 2009. IEEE, 1-5.

KAYSER, H., EWERT, S. D., ANEMÜLLER, J., ROHDENBURG, T., HOHMANN, V. & KOLLMEIER, B. 2009. Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses. *EURASIP Journal on Advances in Signal Processing,* 2009**,** 6.

KITTERICK, P. T., BAILEY, P. J. & SUMMERFIELD, A. Q. 2010. Benefits of knowing who, where, and when in multi-talker listening. *The Journal of the Acoustical Society of America,* 127**,** 2498-2508.

KNUDSEN, L. V., LAPLANTE-LEVESQUE, A., JONES, L., PREMINGER, J. E., NIELSEN, C., LUNNER, T., HICKSON, L., NAYLOR, G. & KRAMER, S. E. 2012. Conducting qualitative research in audiology: A tutorial. *International Journal of Audiology,* 51**,** 83-92.

KOBELT, G. 2013. *Health Economics: An Introduction to Economic Evaluation.*, London: Office of Health Economics.

KOO, T. K. & LI, M. Y. 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine,* 15**,** 155-163.

KUTHUBUTHEEN, J., MITTMANN, N., AMOODI, H., QIAN, W. & CHEN, J. M. 2015. The effect of different utility measures on the cost-effectiveness of bilateral cochlear implantation. *Laryngoscope,* 125**,** 442-7.

LAMBERT, S. D. & LOISELLE, C. G. 2008. Combining individual interviews and focus groups to enhance data richness. *Journal of advanced nursing,* 62**,** 228-237.

LEAVY, P. 2014. *The Oxford handbook of qualitative research*, Oxford library of psychology.

LEFEVER, S., DAL, M., & MATTHIASDOTTIR, A. (2007). Online data collection in academic research: advantages and limitations. *British Journal of Educational Technology*, *38*(4), 574-582.

LEE, T. T., ZIEGLER, J. K., SOMMI, R., SUGAR, C., MAHMOUD, R. & LENERT, L. A. 2000. Comparison of preferences for health outcomes in schizophrenia among stakeholder groups. *Journal of Psychiatric Research,* 34**,** 201-210.

LENERT, L. & KAPLAN, R. M. 2000. Validity and interpretation of preference-based measures of health-related quality of life. *Medical Care,* 38**,** II138-50.

LENERT, L. & STURLEY, A. Acceptability of computerized visual analog scale, time trade-off and standard gamble rating methods in patients and the public. *Proceedings of the American Medical Informatics Association Symposium, 2001. American Medical Informatics Association*, 364.

LENERT, L. A., MICHELSON, D., FLOWERS, C. & BERGEN, M. R. 1995. IMPACT: an object-oriented graphical environment for construction of multimedia patient interviewing software. *Proceedings of Annual Symposium on Computer Application in Medical Care***,** 319-23.

LENERT, L. A. & SOETIKNO, R. M. 1997. Automated computer interviews to elicit utilities: potential applications in the treatment of deep venous thrombosis. *Journal of the American Medical Informatics Association : Journal of American Medical Informatics Association,* 4**,** 49-56.

LENERT, L. A. & STURLEY, A. E. 2002. Use of the internet to study the utility values of the public. *Proceedings of the American Medical Informatics Association Symposium***,** 440-4.

LENERT, L. A., STURLEY, A. P., RAPAPORT, M. H., CHAVEZ, S., MOHR, P. E. & RUPNOW, M. 2004. Public preferences for health states with schizophrenia and a mapping function to estimate utilities from positive and negative symptom scale scores. *Schizophrenia Research,* 71**,** 155-65.

LENERT, L. A., ZIEGLER, J., LEE, T., UNFRED, C. & MAHMOUD, R. 2000. The risks of multimedia methods: effects of actor's race and gender on preferences for health states. *Journal of the American Medical Informatics Association,* 7**,** 177-85.

LLOYD, A., NAFEES, B., NAREWSKA, J., DEWILDE, S. & WATKINS, J. 2006. Health state utilities for metastatic breast cancer. *British Journal of Cancer,* 95**,** 683-90.

LOIZOU, P. C. 2006. Speech processing in vocoder-centric cochlear implants. *Cochlear and brainstem implants.* Karger Publishers.

LONGWORTH, L., YANG, Y., YOUNG, T., MULHERN, B., HERNANDEZ ALAVA, M., MUKURIA, C., ROWEN, D., TOSH, J., TSUCHIYA, A., EVANS, P., DEVIANEE KEETHARUTH, A. & BRAZIER, J. 2014.

Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modelling and survey. *Health Technology Assessment,* 18**,** 1-224.

LOVETT, R. E., KITTERICK, P. T., HEWITT, C. E. & SUMMERFIELD, A. Q. 2010. Bilateral or unilateral cochlear implantation for deaf children: an observational study. *Archives of Disease in Childhood,* 95**,** 107-12.

LTD., Q. I. P. 2014. NVivo qualitative data analysis software; .

LUCAS, L., KATIRI, R. & KITTERICK, P. T. 2018. The psychological and social consequences of single-sided deafness in adulthood. *International Journal of Audiology,* 57**,** 21-30.

LUCAS, L. & KITTERICK, P. T. Describing the health of individuals with a single-sided deafness. *British Academy of Audiology's 11th Annual Conference*., 2014.

MALTERUD, K. 2001. Qualitative research: standards, challenges, and guidelines. *The lancet,* 358**,** 483-488.

MANLY, T., ROBERTSON, I. H., GALLOWAY, M. & HAWKINS, K. 1999. The absent mind: further investigations of sustained attention to response. *Neuropsychologia,* 37**,** 661-70.

MCDOWELL, I. 2006. *Measuring health: a guide to rating scales and questionnaires*, Oxford University Press.

MCLEOD, B., UPFOLD, L. & TAYLOR, A. 2008. Self reported hearing difficulties following excision of vestibular schwannoma. *International journal of audiology,* 47**,** 420-430.

MOORE, T. Voice communication jamming research. Advisory Group for Aerospace Research and Development Conference Proceedings, 1981. DTIC Document.

MORSS, S. E., LENERT, L. A. & FAUSTMAN, W. O. The side effects of antipsychotic drugs and patients' quality of life: patient education and preference assessment with computers and multimedia. *Proceedings of the Annual Symposium on Computer Application in Medical Care, 1993. American Medical Informatics Association, 17.*

MUKAKA, M. M. 2012. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi medical journal : the journal of Medical Association of Malawi,* 24**,** 69-71.

MULHERN, B., BANSBACK, N., BRAZIER, J., BUCKINGHAM, K., CAIRNS, J., DEVLIN, N., DOLAN, P., HOLE, A. R., KAVETSOS, G., LONGWORTH, L., ROWEN, D. & TSUCHIYA, A. 2014. Preparatory study for the revaluation of the EQ-5D tariff: methodology report. *Health Technology Assessment,* 18**,** vii-xxvi, 1-191.

NAFEES, B., STAFFORD, M., GAVRIEL, S., BHALLA, S. & WATKINS, J. 2008. Health state utilities for non small cell lung cancer. *Health and Quality of Life Outcomes,* 6**,** 84.

NEILL, W. T., VALDES, L. A. & TERRY, K. M. 1995. Selective attention and the inhibitory control of cognition. *Interference and inhibition in cognition.* San Diego, CA, US: Academic Press.

NICE. 2013. *Guide to the Methods of Technology Appraisal 2013. NICE article [PMG9].* [Online]. London: National Institute for Health Care Excellence. Available: https://www.nice.org.uk/article/pmg9/chapter/foreword [Accessed].

NICE. 2014. *Who are we.* [Online]. National Institute For Health and Care Excellence. Available: https://www.nice.org.uk/about/who-we-are [Accessed 06/03/2015 2015].

NORMAN, R., CRONIN, P., VINEY, R., KING, M., STREET, D. & RATCLIFFE, J. 2009. International Comparisons in Valuing EQ-5D Health States: A Review and Analysis. *Value in Health,* 12**,** 1194-1200.

NUNNALLY, J. C. 1978. *Psychometric theory.*, New York: McGraw-Hill.

OSBORNE, R. H., DALTON, A., HERTEL, J., SCHROVER, R. & SMITH, D. K. 2012. Health-related quality of life advantage of long-acting injectable antipsychotic treatment for schizophrenia: a time trade-off study. *Health and Quality of Life Outcomes,* 10**,** 35.

PATTON, M. Q. 1990. . *Qualitative Evaluation and Research Methods.* .

PETERSON, N. R., PISONI, D. B. & MIYAMOTO, R. T. 2010. Cochlear implants and spoken language processing abilities: Review and assessment of the literature. *Restorative neurology and neuroscience,* 28**,** 237-250.

PICKARD, A. S. 2015. Is it Time to Update Societal Value Sets for Preference-Based Measures of Health? *PharmacoEconomics,* 33**,** 191-192.

PRINSEN, C. A., VOHRA, S., ROSE, M. R., BOERS, M., TUGWELL, P., CLARKE, M., WILLIAMSON, P. R. & TERWEE, C. B. 2016. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set"–a practical guideline. *Trials,* 17**,** 449.

READ, J. L., QUINN, R. J., BERWICK, D. M., FINEBERG, H. V. & WEINSTEIN, M. C. 1984. Preferences for health outcomes. Comparison of assessment methods. *Medical Decision Making,* 4**,** 315-29.

REASON, J. & LUCAS, D. 1984. Absent-mindedness in shops: its incidence, correlates and consequences. *British Journal of Clinical Psychology,* 23 (Pt 2)**,** 121-31.

ROBINSON, R. 1993. Cost-utility analysis. *British Medical Journal,* 307**,** 859-862.

ROWEN, D., BRAZIER, J., TSUCHIYA, A., YOUNG, T. & IBBOTSON, R. 2012. It's all in the name, or is it? The impact of labeling on health state values. *Medical Decision Making,* 32**,** 31-40.

SANO, H., OKAMOTO, M., OHHASHI, K., INO, T., IWASAKI, S. & OGAWA, K. 2013. Self-reported symptoms in patients with idiopathic sudden sensorineural hearing loss. *Otology & Neurotology,* 34**,** 1405-1410.

SASSI, F. 2006. Calculating QALYs, comparing QALY and DALY calculations. *Health Policy and Planning,* 21**,** 402-408.

SCHOOLER, J. W., REICHLE, E. D. & HALPERN, D. V. 2004. Zoning Out while Reading: Evidence for Dissociations between Experience and Metaconsciousness. *Thinking and seeing: Visual metacognition in adults and children.* Cambridge, MA, US: MIT Press.

SHINGLER, S. L., SWINBURN, P., LLOYD, A., DIAZ, J., ISBELL, R., MANSON, S. & BENSON, C. 2013. Elicitation of health state utilities in soft tissue sarcoma. *Quality of Life Research,* 22**,** 1697-706.

SHOUKRI, M. M., ASYALI, M. H. & DONNER, A. 2004. Sample size requirements for the design of reliability study: review and new results. *Statistical Methods in Medical Research,* 13**,** 251-271.

SHROUT, P. E. & FLEISS, J. L. 1979. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin,* 86**,** 420.

SIMONS, D. J. 2000. Attentional capture and inattentional blindness. *Trends in Cognitive Sciences,* 4**,** 147-155.

SMEDS, K., WOLTERS, F. & RUNG, M. 2015. Estimation of Signal-to-Noise Ratios in Realistic Sound Scenarios. *Journal of the American Academy of Audiology,* 26**,** 183-96.

STALMEIER, P. F., GOLDSTEIN, M. K., HOLMES, A. M., LENERT, L., MIYAMOTO, J., STIGGELBOUT, A. M., TORRANCE, G. W. & TSEVAT, J. 2001. What should be reported in a methods section on utility assessment? *Medical Decision Making,* 21**,** 200-7.

STEIN, K., SUGAR, C., VELIKOVA, G. & STARK, D. 2007. Putting the 'Q' in quality adjusted life years (QALYs) for advanced ovarian cancer - An approach using data clustering methods and the internet. *European Journal of Cancer,* 43**,** 104-13.

STREINER, D. L. & NORMAN, G. R. 2008. *Health measurement scales : a practical guide to their development and use,* Oxford, Oxford University Press.

SUBRAMANIAM, K., EIKELBOOM, R. H., EAGER, K. M. & ATLAS, M. D. 2005. Unilateral profound hearing loss and the effect on quality of life after cerebellopontine angle surgery. *Otolaryngology—Head and Neck Surgery,* 133**,** 339-346.

SUMMERFIELD, A. Q., LOVETT, R. E., BELLENGER, H. & BATTEN, G. 2010. Estimates of the cost-effectiveness of pediatric bilateral cochlear implantation. *Ear and Hearing,* 31**,** 611-24.

SUMMERFIELD, A. Q., MARSHALL, D. H., BARTON, G. R. & BLOOR, K. E. 2002. A cost-utility scenario analysis of bilateral cochlear implantation. *Archives of Otolaryngology–Head & Neck Surgery,* 128**,** 1255-1262.

SUMNER, W., NEASE, R. & LITTENBERG, B. 1991. U-titer: a utility assessment tool. *Proceedings of the Annual Symposium on Computer Application in Medical Care.*, 701-5.

TAYLOR, R. 1990. Interpretation of the Correlation Coefficient: A Basic Review. *Journal of Diagnostic Medical Sonography,* 6**,** 35-39.

TERWEE, C. B., BOT, S. D., DE BOER, M. R., VAN DER WINDT, D. A., KNOL, D. L., DEKKER, J., BOUTER, L. M. & DE VET, H. C. 2007. Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of clinical epidemiology,* 60**,** 34-42.

TORRANCE, G. W. 1976. Social preferences for health states: An empirical evaluation of three measurement techniques. *Socio-Economic Planning Sciences,* 10**,** 129-136.

TORRANCE, G. W. 1986. Measurement of health state utilities for economic appraisal: A review. *Journal of Health Economics,* 5**,** 1-30.

TORRANCE, G. W., THOMAS, W. H. & SACKETT, D. L. 1972. A utility maximization model for evaluation of health care programs. *Health Services Research,* 7**,** 118-33.

UBEL, P. A., LOEWENSTEIN, G. & JEPSON, C. 2003. Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. *Quality of life Research,* 12**,** 599-607.

VAN NOOTEN, F. & CARO, J. J. 2013. Use of relative effectiveness information in reimbursement and pricing decisions in Europe. *Journal of comparative effectiveness research,* 2**,** 33-44.

VAS, V., AKEROYD, M. A. & HALL, D. A. 2017. A Data-Driven Synthesis of Research Evidence for Domains of Hearing Loss, as Reported by Adults With Hearing Loss and Their Communication Partners. *Trends in Hearing,* 21**,** 2331216517734088.

VELLINGA, A., SMIT, J. H., VAN LEEUWEN, E., VAN TILBURG, W. & JONKER, C. 2004. Competence to consent to treatment of geriatric patients: judgements of physicians, family members and the vignette method. *International Journal of Geriatric Psychiatry,* 19**,** 645-654.

VERMEIRE, K. & VAN DE HEYNING, P. 2009. Binaural hearing after cochlear implantation in subjects with unilateral sensorineural deafness and tinnitus. *Audiology and Neurotology,* 14**,** 163-171.

VICEN-BUENO, R., GIL-PITA, R., UTRILLA-MANSO, M. & ALVAREZ-PEREZ, L. A hearing aid simulator to test adaptive signal processing algorithms. 2007 *IEEE International Symposium on Intelligent Signal Processing, 2007. I*EEE, 1-6.

VINCENT, C., ARNDT, S., FIRSZT, J. B., FRAYSSE, B., KITTERICK, P. T., PAPSIN, B. C., SNIK, A., VAN DE HEYNING, P., DEGUINE, O. & MARX, M. 2015. Identification and evaluation of cochlear implant candidates with asymmetrical hearing loss. *Audiology and Neurotology,* 20 Suppl 1**,** 87-9.

WEIR, J. P. 2005. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *The Journal of Strength & Conditioning Research,* 19**,** 231-240.

WELSH, L. W., ROSEN, L. F., WELSH, J. J. & DRAGONETTE, J. E. 2004. Functional impairments due to unilateral deafness. *Annals of Otology, Rhinology & Laryngology,* 113**,** 987-993.

WIE, O. B., HUGO PRIPP, A. & TVETE, O. 2010. Unilateral deafness in adults: effects on communication and social interaction. *Annals of Otology, Rhinology & Laryngology,* 119**,** 772.

WORLD HEALTH ORGANISATION. *Trade, foreign policy, diplomacy and health* [Online]. Available: http://www.who.int/trade/glossary/story046/en/ [Accessed 17/03 2015].

WORLD HEALTH ORGANIZATION. (2017). *Global costs of unaddressed hearing loss and cost-effectiveness of interventions: a WHO report, 2017*. World Health Organization.

WORLD HEALTH ORGANIZATION INTERNATIONAL (2020). *Deafness And Hearing Loss*. [online] Available at: <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss> [Accessed 7 July 2020].

WRIGHT, K. B. 2005. Researching Internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer‐Mediated Communication,* 10**,** 00-00.

YARDLEY, L. 2017. Demonstrating the validity of qualitative research. *The Journal of Positive Psychology,* 12**,** 295-296.