University of Southampton

Faculty of Engineering and Physical Sciences

Institute of Sound and Vibration Research

Auditory fitness for duty: Acoustic stealth awareness

by

Matthew Blyth

Thesis for the degree of Doctorate of Philosophy

September 2019

University of Southampton

Abstract

Faculty of Engineering and Physical Sciences

Institute of Sound and Vibration Research

Thesis for the degree of Doctorate of Philosophy

Auditory fitness for duty: Acoustic stealth awareness

by

Matthew Blyth

Military personnel sometimes have to operate without being detected. In these situations, it is important that the individual has an awareness of their own detectability, both visually and acoustically, in order to operate effectively. Acoustic stealth awareness (ASA) refers to an individual's judgement of their aural detectability with respect to a detector (target). It has been suggested that hearing impairment might affect ASA due to reduced auditory feedback; however, there has been very little research on ASA, including how to measure it and the factors that influence it. Given the potential implications of hearing impairment for an individual's auditory fitness for duty (AFFD), a better understanding of the role of hearing in ASA and how hearing impairment impacts an individual's AFFD is required. The aim of this study was to develop a method for investigating ASA, explore factors that affect judgements and assess the accuracy of judgements.

A number of potential experimental approaches were considered to balance the need to control variables that might influence ASA (background noise, wind noise, etc.) and ecological validity. Experiment 1 investigated egocentric visual distance estimation in virtual reality (VR) and reality in an outdoor open field environment. The results showed that distance estimation was similar on average in the two viewing environments over 25 – 125 m. This suggested that VR could be used in applications where similar distance estimation to reality over these ranges is likely to be important, such as ASA.

Using the same VR environment as for Experiment 1, a novel method for measuring aural detectability judgements was developed. This required subjects to: 1) view a distant target, 2) listen to a sound produced near them, and 3) judge, yes or no, whether the target would be able to detect that sound. Experiment 2 used this developed method to investigate aural detectability judgements for various subject-target distances (25, 50, 100 m) and stimulus types (Gaussian noise, pine cone crunching, whispered digits), measured using normal-hearing civilians. The results showed that judgements were repeatable, sensitive to sound level, sensitive to subject-target distance and dependent on the stimulus type. Experiment 3 measured the absolute thresholds for each of the sounds that people judged in the previous experiment. The results of these two experiments were combined in order to assess the accuracy of aural detectability judgements. In general, people did not make accurate judgements, rather subjects tended to report sounds as undetectable when they probably were detectable. Judgements were found to get less accurate as subject-target distance increased, and were markedly poorer for whispered digits, suggesting that prior experience of sounds might affect judgements. It is concluded

that aural detectability judgements are sensitive to relevant factors such as distance and sound level, but are generally inaccurate, at least for normal-hearing civilians; the degree of error associated with judgements is variable between people, but mostly in the direction that suggests people do not have accurate ASA. This may have implications for the military; further research is required in order to understand if these findings are replicated by military personnel and in real-life acoustic stealth situations, and how hearing impairment affects judgements.

Table of Contents

Table of Tables

Table of Figures

Research Thesis: Declaration of Authorship

Print name:	Matthew Blyth

Title of thesis:	Auditory fitness for duty: Acoustic stealth awareness

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

Signature:		Date:	05.09.2019

Acknowledgements

I would like to acknowledge and thank my supervisor, Dr Daniel Rowan, for his support and guidance throughout the PhD. In particular, his confidence in me, and enthusiasm and passion for science, helped enormously. I also thank my supervisory team, Prof Simon Liversedge, Dr Hannah Semeraro and Dr Adrian Allsopp, for the helpful discussions, support and guidance. I wish to also thank Dr Ben Lineton, Prof Nicci Campbell and Prof Brian Moore for their helpful comments and feedback throughout the assessment process of the PhD. I acknowledge the Royal Centre for Defence Medicine who funded this project.

I also wish to acknowledge Dr Jessica Monaghan, for use of her code and help with programming during the first 18 months of the PhD. I thank Dr Mark Fletcher for use of his Uber-o-meter audiometry code, and helpful discussions (I still won't get in the tank). I acknowledge the Cambridge research group for their loudness model and hearing loss simulation code. I also acknowledge Sam Perry, my neighbour and tutor for all things programming, for his enormous help and patience. Finally, thanks to all the participants that volunteered their time, ears and eyes.

I wish to thank my parents for their love and support, and for valuing my education so highly. I also thank my brother and his wife, Scott and Kyoko, for the invaluable periods of respite in Hong Kong. Finally, I wish to thank my partner, Jasmine, for her love, support, unlimited patience and companionship.

Abbreviations

AFFD	Auditory fitness for duty
NIHL	Noise induced hearing loss
ASA	Acoustic stealth awareness
PTA	Pure tone audiometry
HTL	Hearing threshold level
UK	United Kingdom
SNR	Signal-to-noise ratio
SME	Subject matter expert
SRT	Speech recognition threshold
HPD	Hearing protection devices
VR	Virtual reality
HMD	Head mounted display
ToM	Theory of Mind
PF	Psychometric function
SIN	Speech in noise
REAT	Real ear attenuation at threshold
SPL	Sound pressure level

Chapter 1 Introduction

Hearing is important for most roles within the military. For example, a soldier must use their sense of hearing to communicate, detect enemy threats and maintain situational awareness (Bevis et al., 2014). Failure to complete auditory tasks can have fatal consequences. It is therefore necessary to ensure that those carrying out hearing-critical tasks possess sufficient hearing ability in order for them to do their jobs effectively and safely (Tufts et al., 2009).

Auditory fitness for duty (AFFD) describes the occupational readiness of an individual with regards to their hearing. For most armed forces, including the United Kingdom (UK), personnel regularly have their AFFD assessed. The outcome of this assessment determines whether the individual can stay in their role. If the test fails to detect those who are genuinely unfit for their role because of their hearing, the life of that individual and the lives of those around them are potentially at risk. Conversely, if the test incorrectly determines that an individual is unfit for their role, the military bears an unnecessary economic burden, due to wasted recruits, wasted investment and costly compensation (Yankaskas, 2013). It is, therefore, of great importance that the test used to determine an individual’s AFFD is sufficiently accurate and precise so that the aforementioned errors are minimised.

In a somewhat ironic fashion, those who depend on their hearing the most are typically those at greatest risk of auditory injury. Roles within the military typically involve dangerous levels of noise exposure from machinery, vehicles and weapons (Humes et al., 2005). For example, the standard issue rifle for the British Army (SA80) emits a peak sound level of approximately 160 dB C, measured 0.3 m to the side of the rifle (Paddan and Lower, 2016); unless stood at least 20 m from the rifle, a single gunshot would exceed the upper exposure action value of the *Control of Noise at Work Regulations* (HSE, 2005), necessitating the use of hearing protection. Despite mandatory hearing protection use as part of hearing conservation programmes, many soldiers still do not wear hearing protection, with some citing reasons such as decreased situation awareness and reduced communication abilities (Bevis et al., 2014, Casali et al., 2009). The high noise-exposure and imperfect use of hearing protection leaves military personnel at a heightened risk of noise induced hearing loss (NIHL) relative to most civilian occupations (Yankaskas, 2013).

The UK military uses a hearing conservation programme to monitor the hearing health of its personnel (Defence, 2013c). This involves testing hearing annually, and 6 months prior to and post deployment abroad, using pure tone audiometry (PTA). The results are assessed by medical officers for any deterioration in hearing levels and the individual's AFFD. Therefore, the test that identifies whether an individual possesses sufficient hearing ability for their role is PTA.

The degree to which PTA accurately measures functional hearing has long been questioned (Musiek et al., 2017, Tremblay et al., 2015, Killion and Niquette, 2000, Ferman et al., 1993, Middelweerd et al., 1990). PTA is a measure of hearing threshold levels, which, for diagnostic and rehabilitative purposes, is fundamental to audiology. In this regard, it is useful for monitoring personnel for progression of NIHL, though recent research suggest PTA might not be able to identify some noise-induced cochlear damage (Kujawa and Liberman, 2009, Tepe et al., 2017). For the assessment of AFFD however, a variety of evidence suggests that hearing threshold levels are unlikely to be able to predict functional hearing abilities, especially on an individual basis, as required by the military (Smoorenburg, 1992, Tremblay et al., 2015, Heinrich et al., 2015). An additional but equally important issue that challenges the use of PTA for the assessment of AFFD is the lack of evidence associating job performance to hearing threshold levels. As such, the predictive validity of PTA as a measure of operational effectiveness is lacking.

Recent conflicts have refocussed attention on NIHL in the military (Nelson et al., 2012). Given the impact of the financial and operational implications of inaccurate AFFD assessments, a recent surge in research activity has appeared internationally, motivated by the need for improved hearing conservation and AFFD measures (Phatak et al., 2018, Sheffield et al., 2017, Sheffield et al., 2016a, Tufts et al., 2018). This study focuses on the pursuit of improved AFFD measures and complements other doctoral research activities (Semeraro, 2016) undertaken by this University toward that aim.

This study sought to investigate acoustic stealth awareness (ASA), which describes a person's awareness of their acoustic output in relation to remaining undetected by enemy forces. Acoustic stealth is an important component of a ground close-combat soldier's (e.g. infantry) role; that is, operating (e.g. approaching or observing an enemy) in a manner that avoids detection by audition. Owing to the paucity of evidence for the predictive validity of PTA for functional hearing in military roles, ASA has received very little scientific investigation. Casali et al. (2009) carried out a field study involving teams of soldiers conducting a simulated reconnaissance mission. They explored how hearing protection devices impacted operational effectiveness. They found that hearing protection detrimentally affected stealth, with subject matter experts (SMEs) judging the participants as being too noisy whilst wearing hearing protection. It seems plausible, both from the evidence in the field study, and from analysing the task more generally, that hearing plays an important role in ASA. Therefore, it is possible that hearing loss may impact upon a soldier's ASA, and therefore their AFFD. Research is required in order to better understand ASA and how hearing-impairment impacts upon performance in situations requiring acoustic stealth[1].

This study begins with a review of the literature relevant to AFFD and ASA, and is described in Chapter 2. The review found substantial gaps in knowledge. The only method found in the literature for measuring acoustic stealth was through the use of SMEs, which can only give limited insight towards ASA. It was expected that ASA involves an aural detectability judgment, which involves the individual judging whether or not their acoustic output is aurally detectable to a distant target. The various factors that might affect acoustic stealth, such as how far away the distant target is from the individual, were also discussed. Following the review of the literature and discussion of ASA, the aims of this study were determined as follows:

1. Design suitable methodology for measuring aural detectability judgements
2. Explore the effects of various factors on aural detectability judgements
3. Evaluate the accuracy of aural detectability judgements

Chapter 2 also discusses some challenges associated with AFFD research, especially with the military. These challenges were identified through a previous PhD student's experience. Specifically, the recruitment of hearing impaired military personnel, plus the covariance of age, noise exposure and military service, make designing and conducting AFFD experiments challenging. As such, a series of supplementary experiments, that were undertaken during the early stages of the study, are reported. These experiments explored the use of hearing loss simulations in AFFD research, specifically looking at how the effects of the simulations (Moore et al., 1995, Baer and Moore, 1993) compare to results for real hearing-impaired listeners, and whether a real-time hearing loss simulation device was appropriate. These experiments are summarised in Chapter 2 and reported in more depth in Appendix A.

Chapter 3 describes Experiment 1, which addressed a knowledge gap in visual distance estimation in virtual reality (VR). In order to achieve the first aim of this study, the development of a novel method, the potential use of VR was indicated; however, distance estimation over the range of distances of interest for ASA research has not been

---

[1] How hearing impairment might affect acoustic stealth awareness is exemplified in a 1964 report prepared for the Hearing Sub-Committee of the Royal Naval Personnel Research Committee, in which a hearing-impaired study participant case history was summarised as the following: "Sergeant W.N.N, aged 27. Recently transferred from the weapon-training branch on account of his deafness. Whilst in a night ambush near the Yemen border, he contravened ambush orders which entailed lying down whilst urinating. This caused a noise which he thought to be so soft as to be inaudible to others. It was however heard clearly all around and could easily have given away the position of the ambush. It resulted in a reprimand and subsequently in a medical investigation of his hearing defect which had been brought to notice in this dramatic manner." (Coles, 1964, page 4).

investigated. The results of this study indicated that VR would be appropriate for use in experiments and applications requiring distance estimation to be similar in VR to reality, so the development of a method for measuring aural detectability judgements using VR commenced. The methodological development, including the recording and processing of stimuli and piloting, are described in Chapter 4.

Aural detectability judgements for normal-hearing civilian subjects were measured in Experiment 2, reported in Chapter 5. This experiment measured the effects of subject-target distance, stimulus type, test-retest reliability and stimulus levels of aural detectability judgements. The accuracy of judgements was explored, first by using a model of sound propagation and loudness perception (Moore et al., 2016), and second by behaviourally measuring thresholds for the sounds used in Experiment 2. The behavioural thresholds for the sounds were measured in Experiment 3, and combined with the judgements measured in Experiment 2 in order to assess the accuracy of judgements. Experiment 3 and the assessment of accuracy of judgments is reported in Chapter 6. A general discussion of ASA, implications for AFFD and suggested future work is given in Chapter 7.

The primary contribution to knowledge made by this thesis is a better understanding of ASA and aural detectability judgements. Using a novel method, this study shows that normal-hearing civilians make judgements that, at least in the situations simulated here, are generally not accurate, where they judge themselves to be undetectable when it is probable that they are. Factors such as the sound level, and how far away the target is, are factored into judgements, but not necessarily accurately. Judgements seem to be affected by non-auditory factors; for example, judgements for whispers were markedly less accurate than judgements for Gaussian noise and pine cones, suggesting that previous experience with stimuli biased judgements. These findings may have implications for the military, though more research is required to explore ASA and aural detectability judgements for normal-hearing and hearing-impaired military personnel; this study contributed a novel method that could be used to conduct this further research. Other contributions include a better understanding of distance estimation in VR over distances previously not investigated, and various hearing loss simulations in relation to AFFD research. Distance estimates were similar in VR to reality over 25 – 125 m, in an outdoor environment where similar visual cues were available in the two environments. Hearing loss simulations gave similar effects on the identification of speech in noise to real hearing-impaired listeners, providing a distorting and audibility component of the simulation was included. The effect of the loudness recruitment simulation only affected speech in noise performance when the speech was at low sensation levels.

Aspects of this study have been reported at a number of conferences:

Blyth M, Monaghan J, Rowan D. Hearing impairment simulation: Exploring the effects of threshold elevation, loudness recruitment and spectral smearing on listening to speech in noise. Poster presented at the British Academy of Audiology Annual Conference; 2015 Nov 26th – 27th; Harrogate, UK

Blyth M, Monaghan J, Rowan D. Development of a real-time hearing impairment simulation using an iPhone application. Poster presented at the British Society of Audiology Annual Conference; 2016, 25 – 27th April; Coventry, UK

Blyth M, Rowan D, Liversedge S, Allsopp A. Development of a novel method for investigating acoustic stealth awareness. Poster presented at the British Society of Audiology Annual Conference; 2017, 29th – 30th June; Harrogate, UK

Blyth M, Semeraro H, Rowan D, Allsopp A. Do you know when you're being too loud? Aural detectability judgements in normal-hearing civilians. Poster presentation at the British Society of Audiology Basic Auditory Science meeting; 2018, 3rd – 4th Sept; Newcastle, UK

Chapter 2 Background

2.1 Overview

This chapter describes the background to the thesis and the context to which this work contributes. First, the concept of auditory fitness for duty (AFFD) is described, followed by a discussion of the current procedures employed by the UK armed forces. The approach taken locally and internationally to improve AFFD procedures is then described, including the challenges involved in researching this area and population. Acoustic stealth awareness (ASA), the topic of this thesis, is introduced and highlighted as a vastly understudied area of AFFD.

2.2 Auditory fitness for duty

2.2.1 Introduction

Occupational health assessments are carried out in a variety of occupations to ensure that employees are able to perform their jobs safely and effectively (Tufts et al., 2009). For instance, strength tests in the United Kingdom (UK) Royal Navy ensure that personnel can cope with the intense physical demands of the job (Bilzon et al., 2002). Auditory fitness for duty (AFFD) testing refers to the occupational health assessment of hearing and determines if an individual possesses sufficient hearing ability to carry out their job safely and effectively (Tufts et al., 2009, Tufts et al., 2018, Semeraro, 2016, Brungart, 2014).

AFFD testing is usually required for occupations that have hearing-critical tasks (Tufts et al., 2009). This means that there are tasks associated with the job role that depend on the sense of hearing and that there are consequences if the task is not performed successfully (Laroche et al., 2008, Semeraro et al., 2015, Bevis et al., 2014). Occupations that have hearing-critical tasks and require AFFD testing include the military, firefighting and the police. For example, if a firefighter is unable to communicate in a critical situation due to an unmanaged hearing impairment then firefighters', and civilians', lives could be at risk; it is therefore important that firefighters are tested to ensure that they are able to communicate whilst performing their job.

AFFD testing usually involves comparing an individual's hearing test result to job-specific criteria to determine the individual's fitness for duty (Tufts et al., 2009). Most AFFD standards are based on Pure Tone Audiometry (PTA), a tone detection in quiet test that measures hearing threshold levels (HTLs) over a range of frequencies (Defence, 2013b).

The PTA result is compared to occupation-specific HTL criteria and the individual's fitness for duty is determined. For example, Michigan State Police require their officers to be able to hear pure tones with levels below than 25 dB HL from 500 to 6000 Hz (MCOLES, 2016).

It is widely accepted that AFFD testing is required for the military profession (Tufts et al., 2009). It is known that the sense of hearing is critical for job performance and safety, as shown in focus group studies (Semeraro et al., 2015, Bevis et al., 2014) and behavioural studies showing the impact of hearing impairment on military-specific hearing tasks and combat operations (Sheffield et al., 2016a, Sheffield et al., 2015, Brungart, 2014, Semeraro, 2016, Garinther et al., 1995, Peters and Garinther, 1990). Moreover, the military profession typically involves being exposed to dangerously high sound levels which pose a significant risk of sensorineural hearing loss, typically noise induced hearing loss[2] (Biggs and Everest, 2011, Humes et al., 2005, Powell and Forrest, 1988, Tanenholtz, 1968). The heightened risk of hearing loss in the military compared to civilian populations places a duty of care on the employers to protect and conserve the employee's hearing, meaning that hearing testing is also required for hearing conservation purposes (HSE, 2005, Humes et al., 2005).

AFFD testing and hearing conservation procedures for military personnel in most countries (e.g. US, Canada, UK) are based on PTA (Tufts et al., 2009). For a variety of reasons, we argue here that a PTA-based testing protocol is unlikely to be fit for purpose and that improvements in the AFFD protocol are required. The following sections outline and critique the current protocol employed by the UK Armed Forces, and discuss the research objectives required to improve AFFD testing.

2.2.2 Auditory fitness for duty testing in the UK Armed Forces

The UK Armed Forces employ the PULHHEEMS system of medical classification to determine an individual's physical and mental health status, and consequently their deployability (fitness for duty) (Fletcher, 1949, Rona et al., 2006, Defence, 2013c). Each letter of PULHHEEMS represents a different aspect of an individual's health: Physical, Upper limbs, Lower limbs, Hearing (left and right), Eyesight (left and right), Mental capacity and emotional Stability. Each aspect has its own assessment protocol and is given a rating of 1 through to 8. A score of 1 represents the highest possible ability in that dimension, with 8 representing the lowest ability. The PULHHEEMS system is a simple, practical and effective approach to monitoring the employability and health of a vast

---

[2] For a detailed review of noise exposure in the military and the risk of hearing loss and tinnitus, see Humes et al. (2005).

number of employees, but it has been criticised due to occasional inappropriate deployment of personnel (Hodgetts and Greasley, 2003) and recruit wastage (Jefferson, 1990). Errors in the occupational health assessment of an individual can lead to safety issues and financial losses for the Ministry of Defence; there is therefore clear motivation for ensuring the most accurate assessment of fitness for duty possible (Tufts et al., 2010).

The hearing assessment component of the medical assessment of personnel involves automated PTA at 0.5, 1, 2, 3, 4 and 6 kHz carried out annually, according to the Joint Service Protocol (JSP) 950 (Defence, 2013c). The sums of the HTLs for low frequencies (0.5, 1, 2 kHz) and for high frequencies (3, 4, 6 kHz) are calculated, and the hearing acuity is assigned a 'H' category. The H categories are shown in Table 2-1.

Table 2-1 - JSP 950 assessment of audiograms. H categories are assigned according to the sum of hearing threshold levels.

H Category	**Sum of hearing level at low frequencies (dB HL)**	**Sum of hearing level at high frequencies (dB HL)**	**General Description**
1	Not more than 45 (No single level to be more than 20 dB)	Not more than 45 (Level not to be more than 30 dB at 6 kHz or 20 dB at any other frequency)	Good hearing
2	Not more than 84	Not more than 123	Acceptable practical hearing for service purposes.
3	Not more than 150	Not more than 210	Impaired hearing. The hearing level at which most personnel are unfit for entry to service.
4	More than 150	More than 210	Very poor hearing. Below entry standard for the Services.
8	More than 150	More than 210	Medically unfit for service.

Following the identification of H categories, the individual is assessed for (Defence, 2013a):

- *AFFD.* Does H category match the required category for the role? If no, refer to occupational health or approved audiology service for full assessment.
- *Rapid hearing loss.* Has there been a loss of 30 dB or more at high frequencies?

- *Asymmetry*. Is there a difference between the sum of 1, 2, 3 and 4 kHz of 45 dB or more in each ear?
- *Compare to normal values for age/gender*. Is the individual's hearing deteriorating faster than expected?

These assessments aim to identify those who are unlikely to be fit for their current role (AFFD) assessment and identify those at risk of, or showing signs of, hearing loss (hearing conservation). The scope of this project is to critique the AFFD assessment portion of the protocol, not the hearing conservation aspect. It should be noted that it is assumed here that the automated PTA testing is carried out with full compliance to the requirements and conditions described in JSP 950, including ambient noise levels, training requirements of staff and conditions under which testing takes place. The following discussion is conducted under these assumptions.

2.2.3 Is the AFFD protocol fit for duty?

The AFFD assessment employed by the UK Armed Forces ultimately rests on the sensitivity and specificity of automated PTA to identify those who do not have sufficient hearing abilities to carry out their job safely and effectively. We argue that the use of PTA is unlikely to be fit for purpose, for two main reasons. First, PTA is a measure of hearing acuity, not functional hearing ability (e.g. comprehending speech in complex acoustic environments, sound localisation, sound identification); the ability of PTA to measure hearing function therefore depends on the strength of the association between PTA and functional hearing ability. Second, the paucity of evidence associating specific job roles to required hearing threshold levels renders the predictive validity of PTA as a measure of AFFD unknown.

Many studies show an association between hearing acuity (measured by PTA) and a variety of listening tasks (Smoorenburg, 1992, Heinrich et al., 2015, Picard et al., 1999, Engelberg, 1965). This is partly explained by the role of audibility in functional hearing; a listener's hearing acuity needs to be sufficient that signal of interest can be accessed (Pavlovic, 2006). However, hearing acuity alone cannot predict functional hearing on an individual basis, especially in noisy environments and for listeners with sensorineural hearing loss (Tremblay et al., 2015, Killion and Niquette, 2000, Plack et al., 2014, Musiek et al., 2017, Ferman et al., 1993, Middelweerd et al., 1990, Phatak et al., 2018). This widely accepted view is based on the involvement of additional factors when listening to complex sounds in complex environments (Plomp, 1986, Houtgast and Festen, 2008, Surprenant and Watson, 2001). These factors include temporal processing, frequency selectivity, intensity discrimination and cognitive abilities (Phillips, 1999, Schoof and Rosen, 2014, Houtgast and Festen, 2008). PTA is also not an optimal test for the identification of other factors affecting functional hearing, including cochlear dead regions (Moore, 2004) and cochlear synaptopathy (Plack et al., 2016, Kujawa and Liberman, 2009, Tepe et al., 2017).

The correlations between PTA thresholds and speech in noise thresholds range from 0.4 to 0.9 depending on the task and population (Harris, 1965, Humes and Roberts, 1990, Smoorenburg, 1992, Kramer et al., 1996, Heinrich et al., 2015, Engelberg, 1965). This shows that PTA is a good predictor of speech-in-noise performance on average, but it cannot explain all the variance observed in speech-in-noise performance. Given that an AFFD testing procedure makes decisions about an individual's fitness for duty, it is possible that some individuals who do possess the functional hearing ability required to perform the job are being removed from service unnecessarily, and individuals that do not possess sufficient functional hearing but have HTLs below the criterion H categories are being inappropriately deployed.

Further to the argument that PTA cannot predict functional hearing on an individual basis, the evidence base supporting why each job role requires each H category is limited. Due to this paucity of evidence, it is not possible to determine whether each H category is suitable for each specific job role. The use of PTA as a test of AFFD is therefore lacking predictive validity (the ability of a test to predict an outcome measure, e.g. the fitness for duty). The somewhat vague definitions of functional hearing associated with each H category (e.g. 'H1 - Good hearing', 'H2 - Acceptable practical hearing for service purposes') are representative of the correlations between PTA and functional hearing tests. However, given the potential safety and financial implications of incorrectly determining an *individual's* fitness for duty, we argue that more evidence associating the functional hearing requirements of specific job roles to AFFD measures is needed in order to improve the predictive validity of PTA as a measure of AFFD.

Over the past decade, researchers have attempted to better understand the link between hearing impairment and performance on specific military tasks (Brungart, 2014, Semeraro, 2016, Sheffield et al., 2016a, Sheffield et al., 2016b, Sheffield et al., 2015, Mentel et al., 2013). This research usually involves creating simulations of military tasks where hearing is important and measuring the effect of hearing loss on task performance. These studies differentiate themselves from other non-military studies investigating the impact of hearing loss by using outcome measures that are more relevant to the overall task performance. This is important, as troops may utilise compensation strategies (increased use of visual cues, exercising more caution) to overcome the challenges imposed by hearing impairment (Keller et al., 2017), which traditional measures (e.g. percentage of correctly identified words) might not capture.

Sheffield et al. (2016b) carried out a study where teams of soldiers battled to achieve mission objectives, using real weapons with blank ammunition. Each team member was GPS tracked and had their lethality (number of enemies eliminated), survivability (number of surviving team members) and other measures (mission progress, friendly fire incidence, etc.) tracked. Each group was given a different level of simulated hearing loss, through use of a real-time hearing loss simulation device. Each team completed the mission several times, each experiencing different hearing profiles in a balanced order. Each team was assigned a composite score of their performance based on the various outcome measures. The results showed a clear negative impact of hearing loss on composite score. Interestingly, the teams with the most experienced members (in terms of service) benefitted more from having good hearing acuity, which suggests that the more experienced team used communication more effectively. Perhaps most interestingly, this study showed clear evidence of a measurable decline in performance when teams had a mild hearing loss, equivalent to a H2 hearing category. This level of hearing acuity is considered acceptable for close-combat (front-line) soldiers in the UK and US military, suggesting that individuals with compromising levels of hearing loss may be considered fit for duty.

In a study looking specifically at the effect of speech intelligibility on performance in a US Navy command and control centre, Keller et al. (2017) simulated a task in which US Navy personnel had to control a naval battle situation. The task involved viewing several screens with dynamic information, receiving and communicating information via a headset from several channels and responding appropriately as the scenario progressed. Speech intelligibility was controlled using real-time speech intelligibility software, which used an automated gain control circuit to maintain a consistent signal level and mixed the speech with white noise to maintain fixed SNRs throughout the simulation. Task performance was measured by eye movements (duration of gaze toward relevant stimuli) and questionnaires filled out by the participants and observing subject matter experts (SMEs). The results showed a systematic degradation in performance with decreasing speech intelligibility on all measures; gaze time on irrelevant stimuli, requests for repeat-backs and failure to respond to messages all increased with decreasing speech intelligibility. Self-reported difficulty increased and SME judged that performance decreased with decreasing speech intelligibility. The authors suggested that the eye movement data indicated that the participants employed compensatory strategies (requested repeat-backs, diverted visual attention) to overcome the speech intelligibility difficulties, but these were not sufficient to overcome the challenges presented by degraded speech intelligibility, most markedly in the poorest speech intelligibility conditions. An interesting finding was that performance degraded slightly between 100% and 80% intelligibility, and degraded markedly between the 80% and 60% intelligibility. This finding illustrates that an

AFFD criterion could be set based on speech intelligibility, whereby personnel must be able to achieve a minimum speech intelligibility score in order to be suitable for that role.

These studies have been highly useful in exploring the impact of hearing impairment on military task performance, as they have highlighted the multi-faceted degradation in performance that traditional approaches (e.g. effect of hearing loss on communication) fail to measure. A limitation to this approach, however, is that the data are specific to the simulated task and therefore are restricted in their generalisability. More research is required to gain a comprehensive understanding of how hearing loss can impact on performance across a wider range of situations experienced by military personnel. Once a more comprehensive understanding of how hearing loss impacts task performance is achieved, measures of AFFD with high predictive validity can be developed.

In summary, PTA is unlikely to be a fully satisfactory measure of AFFD due to: 1) the inadequate association between HTLs and functional hearing, and 2) the lack of evidence of the predictive validity of PTA as an AFFD test for specific job roles. Recent studies have begun to shed light on how hearing loss affects performance on a selection of military tasks. A deeper and more comprehensive understanding of how hearing loss affects military task performance is required in order to develop measures of AFFD with high predictive validity.

2.2.4 Challenges in AFFD research

A significant challenge that must be addressed when conducting AFFD research is the covariance of military experience and hearing loss (Helfer et al., 2010, Trost and Shaw, 2007, Rovig et al., 2004). As an individual soldier's experience on the battlefield increases, so does his/her noise exposure and age. It therefore becomes difficult to disentangle the effects of experience and hearing loss (related to age or noise exposure) on task performance (Brungart, 2014).

One approach employed by AFFD researchers is the use of hearing loss simulations (Sheffield et al., 2016a, Sheffield et al., 2016b, Mentel et al., 2013). By using hearing loss simulations, a repeated measures experimental design can be used, which overcomes the confounding effects of experience by measuring performance on a task using the same normal-hearing participant in a variety of hearing loss conditions. A major limitation of this approach is that the results of the experiment depend on the hearing loss simulation; any inaccuracies associated with the simulations, therefore, become inherent limitations of the experiment. If the data are to be used by stakeholders and in policy making, there are potential issues with the external validity of the findings.

An additional significant challenge in conducting AFFD research, especially for civilian researchers, is the recruitment of military and hearing-impaired volunteers (Semeraro, 2016). Previous work by our group found substantial barriers to sample recruitment in the military, at least for civilian researchers. Firstly, military personnel follow tight timetables and have little or no time available for experiment participation. They also often do not wish to volunteer their free time by taking part in experiments. Secondly, hearing-impaired personnel are identified via the Defence Audiology Service and can be accessed for experimental trials at this location. The location of the service (south coast of UK) means that patients can travel long distances in order to attend appointments and therefore are often reluctant to take part in studies due to extensive travel time and wanting to rush home. Thirdly, experiments involving military personnel are subject to a robust ethical approval procedure, which ensures that volunteers are participating voluntarily. The hierarchical structure of the military meant that when senior members offered access to lower ranking personnel, it was not always apparent that those lower ranking personnel were attending entirely voluntarily. This poses ethical challenges. Finally, civilian researchers working with the military depend on contacts within the military to achieve the logistical procedures of experiments; this was found to be particularly challenging. A potential, albeit imperfect, solution to these recruitment challenges is to use normal-hearing military personnel with hearing loss simulations. This increases the pool of potential volunteers.

2.2.5 Supplementary experiments: Hearing loss simulation for AFFD research purposes

During the early stages of the PhD, the potential use of hearing loss simulations in order to overcome some of the previously discussed challenges of recruitment and experimental design (Section 2.2.4) was explored. As aforementioned, experiments that explore hearing loss by use of hearing loss simulation are dependent on the validity of the simulation. The limited validation of hearing loss simulators for use in AFFD research motivated a series of experiments which:

1. Compared speech-in-noise speech recognition thresholds (SRTs) for simulated and real hearing-impaired listeners
2. Further explored the effect of presentation level on speech-in-noise SRTs.
3. Investigated the feasibility of a real-time hearing loss simulator by measuring its real-ear-attenuation thresholds (REAT) and latency.

The findings of these experiments are summarised here and reported in greater detail in Appendix A.

2.2.5.1 *Experiment S1: A comparison of the effects of real and simulated hearing loss on speech in noise.*

Experiment S1 (Section A.2) compared the effects of various hearing loss simulations on the monaural SRT for BKB sentences in stationary noise, in a variety of processing conditions. The simulations used were designed to simulate threshold elevation by frequency-specific attenuation, loudness recruitment by envelope expansion (Moore and Glasberg, 1993) and reduced frequency selectivity by spectral smearing (Baer and Moore, 1993). The SRTs were compared to results for a hearing-impaired sample from a previously collected dataset.

The results showed that threshold elevation and spectral smearing negatively affected the SRT, with the magnitude of effect depending on the audibility of the target speech. The loudness recruitment simulation had a beneficial effect, where the SRT improved relative to the threshold elevation only conditions. This finding motivated Experiment S2.

When comparing the simulated hearing loss results to the mean SRT for hearing-impaired listeners, the hearing loss simulations were shown to affect the SRT in stationary noise by a comparable amount.

2.2.5.2 *Experiment S2: Further exploration of the effects of loudness recruitment on speech in noise.*

Following the results of Experiment S1, the apparent beneficial effect of loudness recruitment was explored (Section 0). Relative to the threshold elevation condition, for an identical input, the loudness recruitment simulation would give an increased output. For low sensation levels, this could improve the audibility of the target speech, and therefore the SRT. Four conditions were included to test this hypothesis: two presentation levels (one at a low sensation level that affected the audibility of the target speech, and one 15 dB higher) and two processing condition (threshold elevation or loudness recruitment). If the increased audibility due to the simulation explained the beneficial effect, an improvement in SRT would only occur at the lower presentation level. An improvement in SRT was measured at the lower presentation level, but not at the higher presentation level. This confirmed that the beneficial effect was related to the increased audibility in the loudness recruitment simulation at low sensation levels.

The combined results of experiment S1 and S2 suggest that the use of hearing loss simulations can produce similar effects on speech in noise performance to true hearing impairment. This can be achieved by simulating the auditory and distortion components of hearing loss. It is recommended that spectral smearing and envelope expansion are used to model the distortion component of sensorineural hearing loss with a threshold elevation simulation to model the audibility components.

2.2.5.3 *Experiment S3: Feasibility study of a real-time hearing loss simulation.*

The third supplementary experiment investigated two potential limitations of a real-time hearing loss simulation device, that was developed at the University of Southampton by Dr Jessica Monaghan. Firstly, the latency of the device was shown to be 18 – 24 msec, depending on the processing condition. Secondly, the real-ear attenuation at threshold (REAT) was measured in order to determine the maximum magnitude of hearing loss that could be simulated (by attenuating the direct path of sound). The results showed that the maximum hearing loss that could be simulated was approximately 30 dB HL at low frequencies and 45 dB HL in the high frequencies. It was decided that this was insufficient for AFFD research, as H3 hearing loss profiles are of interest, as these are considered un-deployable by the military (Section 2.2.2).

2.2.5.4 *Summary of hearing loss simulation experiments and implications for AFFD research*

The experiments contributed useful evidence concerning whether hearing loss simulations are appropriate for AFFD research. The results suggest that for 'offline' listening tasks that take place under laboratory conditions with stimuli presented through headphones, hearing loss simulations can reproduce the effects of hearing loss on speech performance in noise well. If a hearing loss simulation is required for field studies, the current device is not suitable, and further development is required.

It should be noted that hearing loss simulations ultimately cannot substitute for real hearing-impaired listeners, as there are effects that a simulation will never be able to reproduce. The use of hearing loss simulations can only alter bottom-up aspects of hearing by manipulating the input to the system; it cannot, by virtue of the technique, manipulate the top-down processes of hearing. This would be valid if hearing were entirely a bottom-up process; this is not the case, as top-down processes are crucial for hearing, especially in complex environments (Shinn-Cunningham et al., 2005, Heinrich et al., 2015, Zekveld et al., 2011, Ronnberg et al., 2010, Shinn-Cunningham, 2008). In conclusion, for establishing an evidence base for policy-making decisions, it is recommended that evidence from both hearing loss simulations and true hearing-impaired listeners is used.

2.2.6 Improving auditory fitness for duty measures

In recent decades there have been efforts to improve measures of AFFD for various occupations, including the UK Armed Forces (Semeraro, 2016), the Department of Fisheries and Ocean Canada (Giguère et al., 2003) and the Canadian Coast Guard (Giguere et al., 2008). All research efforts have the same goal of developing measures of functional hearing with improved predictive validity that can replace, or compliment, PTA (Soli et al., 2018a).

Various approaches have been adopted in order to achieve this shared goal. Laroche et al. (2003) first identified the hearing requirements and hearing tasks of the occupation and recorded the background noise that these tasks occur in. Next, they identified already-validated clinical measures of functional hearing that can assess the auditory demands of the job found in the first stage (e.g. the Hearing in noise test (HINT)). The third step involved statistically relating the screening measures (HINT) to performance in real-world noise environments using normal-hearing listeners; the performance in real-world noise environments was measured by playing the HINT sentences in a variety of the recorded background noises. A statistical model relating the screening measure score to the real-world noise environment score was created. This statistical model was than validated using a different sample of listeners, and the predictive validity was assessed using hearing-impaired listeners. The statistical model was then used to create screening criteria in order to identify listeners who would be able perform the hearing tasks associated with the job. This process resulted in a functional screening test (HINT) with pass criteria based on the specific requirements of the job.

A more objective approach, which involved less laborious validation, based on predicting speech intelligibility in real-world environments, was employed by Soli et al. (2018a). They used the individual's PTA thresholds, SRT and the Extended Speech Intelligibility Index (ESII) values to predict the likelihood of effective communication where hearing tasks need to be performed. The ESII was developed and validated in a companion paper (Soli et al., 2018b). The output of the procedure is a model for defining the functional requirements of the job and a means to ensure an individual can perform at a sufficient level. Soli et al. (2018a) argue this approach provides an objective rationale for defining the criteria used to determine AFFD.

In summary, various approaches have been used to improve the predictive validity of AFFD measures and define suitable occupation-specific screening criteria. Their rationales are focussed on using screening tests that are based on the job and therefore already present an improvement on the PTA-based H categories currently used by many occupations. They generally explore the statistical link between performance in the real-world situation and a clinical screening test, and provide screening criteria in order to determine fitness for duty. Specific evidence of the effectiveness of an updated protocol versus the previous is, so far, lacking, so it is not possible to infer whether the new AFFD protocols are more accurate.

2.2.7 Hearing tasks in the military are not limited to speech communication

The significant flaw in the aforementioned approaches to improving AFFD measures is the assumption that the only task an individual uses their hearing for is speech communication. For military personnel, particularly those engaging in ground close-combat (e.g. infantrymen), this is not the case (Scharine et al., 2009). Our research group adopted a similar approach to Laroche et al. (2003) and first identified the hearing tasks that military personnel regularly perform and reasons why performance might be reduced (Bevis et al., 2014). This was carried out by forming focus groups with infantrymen with varying lengths of service. A summary of the themes and infantry auditory tasks that emerged from the focus groups is shown in Table 2-2. A follow up questionnaire-based study measured the frequency and cost of poor task performance of the auditory tasks identified in the focus group study. This identified the mission-critical auditory tasks (MCATs) associated with infantry job performance (Semeraro et al., 2015). The focus group approach of job analysis is limited by the experiences of the sample. In a review of the validity and reliability of physical employment standards, Milligan et al. (2016) argue that subject matter experts (SMEs) should form the panel used to identify critical tasks associated with the job. Whilst the participants in our focus groups were selected in order to gather the ideas and opinions of a range of personnel, only 19% of the sample had completed more than two tours. As a result, it is possible that the identified tasks are not fully representative of the hearing tasks associated with the job. As discussed in Section 2.2.4, accessing military participants is highly challenging, so this critical analysis of the study is given in light of those recruitment challenges.

Table 2-2 - Themes, subthemes and examples identified in a focus group study of the hearing requirements of infantrymen (Bevis et al., 2014).

Theme	**Subtheme**	**Example**
Auditory tasks	Sound detection	Detecting a malfunction in an item of machinery
	Speech communication	Hearing directions on patrol
	Sound localisation	Locating the moving sound source of footsteps
Reasons for reduced performance	Background noise	*"There will be an engine running or something all the time"*
	Hearing protection devices	*"So you won't know what is going on around you, you can't hearing nothing"*
	Stress	*"In a contact when every ones flapping, but you're running over there because you can't hear and you're headless chicken and you're scared"*
	Attention difficulties	*"You can't focus on what you are meant to be hearing, it is nothing to do with your hearing, you just can't process everything; take everything in, do you see what I mean? You can't do it."*

The focus group and questionnaire studies provided clear evidence that the auditory tasks associated with close combat comprise more than just speech communication. AFFD measures for military personnel should therefore consider the predictive validity of a screening test for tasks in addition to speech communication. Our research group's approach to improving AFFD measures has therefore been to consider the effect of hearing loss on a more comprehensive range of hearing tasks.

2.2.8 Acoustic stealth awareness

An aspect of operational effectiveness for military personnel in which hearing is thought to play a crucial role is acoustic stealth awareness (ASA) (Sheffield et al., 2017, Esquivel et al., 2018, Casali et al., 2009). This refers to situations where soldiers must operate stealthily in order to avoid being aurally detected by nearby enemy. The consequences of having poor acoustic stealth (e.g. stomping around creating high noise levels and increasing their aural detectability) is detection by enemy forces, which could lead to operational failure and fatalities (Casali et al., 2009). It is therefore critical that soldiers have an awareness of their acoustic stealth in order to operate effectively and safely (e.g. remain undetected, seek cover and engage if detected by the enemy).

The US military use an “Aural non-detectability” standard to ensure that new equipment is suitably quiet (Garinther et al., 1985). This standard states the permissible sound levels that a piece of equipment can emit in order for it to be sufficiently aurally non-detectable. The standard was developed using a two-stage model of aural detectability. The first stage used a sound propagation model that calculated the attenuation between a source and receiver that are some distance apart, and included the inverse square law (6 dB loss per doubling of distance), atmospheric absorption, ground effect and other meteorological effects. The second stage modelled the auditory perception of the sound following propagation (the sound at the receiver location) to determine if it could be detected. The standard gives the maximum permissible sound level at a reference distance of 1 m from the source for each 1/3 octave band, for that sound source to be non-detectable, for detection distances between 10 and 6000 m. The standard helps researcher and development teams to measure whether a new piece of equipment is suitably non-detectable for the likely situations it will be used in. A recent example of this is the testing of a new set of night-vision goggles (Gaston et al., 2013); these were found to be non-detectable up to a distance of around 10 m, meaning a normal-hearing listener needs to be 10 m or closer to be able to aurally detect the night-vision goggles. Whilst useful, this standard only refers to the objective limits equipment can have; it does not relate to a soldier’s awareness of their aural detectability.

There is a severe lack of published studies investigating ASA. A thorough literature search revealed only one study that did included a measure of acoustic stealth. This study investigated the effects of hearing protection devices (HPDs) on a simulated battlefield task (Casali et al., 2009). Teams of soldiers had to complete a reconnaissance-style mission by gathering information about an enemy camp and returning back to a check point, whilst navigating various threats and avoiding detection. The purpose of the study was to assess how various HPDs affected performance. Three HPDs were used: passive foam plugs and two different active devices. As well as objective measures, such as time taken and the distance from threats at which they identified the threats, the outcome measures included observations by SMEs, including their acoustic stealth, who wrote comments about their performance. They also asked the soldiers after completing the mission to rate how the HPD ‘interfered with stealth’, as well as other questions related to the mission, on a Likert-style rating scale. The results showed many comments from SMEs indicating that the soldiers performed with poor acoustic stealth. The comments included:

- "would have been heard had the enemy been real, and not simulated”
- “the Squad continued to talk louder than normal”
- "too many branches were crunching"
- "Squad members spoke too loud continuously"
- "at one point, [a Squad member] spoke loud when it [would have been] appropriate to whisper”
- "movement was loud, people didn't consider the noise [that] they were making"
- "Squad was very loud in their footsteps"

The post-mission soldier ratings found that all HPDs used interfered with stealth, with the passive foam plugs having the greatest negative impact.

Whilst this study does little to improve the overall understanding of ASA, it does highlight that impaired auditory feedback (HPDs) affected their acoustic stealth and their ASA. The results are difficult to interpret due to the subjective nature of the outcome measures used. At no point did the authors objectively quantify the aural detectability of the soldiers’ movements, so it is not possible to know how effective their ASA was. This study was useful in that it measured acoustic stealth in a real-world battlefield simulation, but it highlighted the need for improved measures and understanding of acoustic stealth.

2.2.9 Gap in knowledge: acoustic stealth awareness

Following the research effort for improved AFFD measures from, and communication with, various groups internationally, a substantial gap in knowledge has been identified for ASA. Hearing is thought to be important for ASA because, presumably, auditory feedback of one’s own acoustic output is necessary for judging whether or not one’s acoustic stealth is compromised. If this presumption is correct, there are implications for an individual’s AFFD, as hearing loss would remove or reduce the available auditory feedback and result in reduced operational effectiveness and safety. There is, therefore, clear motivation for investigating ASA and the impact of hearing loss. These investigations should aim to:

1. Understand the role of hearing in acoustic stealth awareness
2. Determine the effects of hearing loss upon performance in tasks requiring acoustic stealth.

If hearing loss negatively impacts performance on tasks requiring acoustic stealth awareness, measures of auditory function should be related to performance on acoustic stealth tasks in order to determine AFFD standards.

Owing to the substantial knowledge gap in the area of ASA, a program of research is required to achieve the aims outlined above. The primary aim of this thesis is to increase the knowledge of ASA (aim 1, above) in order to make progress towards achieving the aims outlined above. The following sections discuss ASA in detail, including the factors that affect acoustic stealth and the methodological approach adopted in this thesis.

2.2.10 Summary of auditory fitness for duty

This discussion of AFFD in the military identified the following:

- AFFD testing is required for occupations where the ability to hear is critical for effective and safe job performance. Inaccurate AFFD protocols can lead to wasted resources and safety concerns.
- The UK Armed Forces utilise an AFFD protocol base on PTA. Based on the known inability of PTA to predict functional hearing ability on an individual basis, and the lack of evidence for the predictive validity of PTA, improvements are required to this protocol.
- AFFD research is challenging for methodological and recruitment reasons. A series of experiments was conducted in order to further investigate whether hearing loss simulations could produce similar effects on speech in noise performance to real hearing loss, and if a recently developed real-time hearing-loss simulation device could be used. The results showed that the real-time device could not simulate hearing losses of sufficient magnitude due to insufficient attenuation of direct-path sound, but the ‘offline’ hearing loss simulations replicated the effect of real hearing loss on speech in noise performance.
- Recent AFFD research has used speech-based screening tests as measures of AFFD. This approach excludes the impact of hearing loss on tasks that do not require speech perception.

- Acoustic stealth awareness has been identified as an area of AFFD requiring research. The aims of the research should be to understand the role of hearing and the effect of hearing loss in acoustic stealth awareness. If hearing loss negatively impacts performance, measures of auditory function should be identified as part of an AFFD protocol.
- The primary aim of this thesis is to increase the understanding of acoustic stealth awareness.

2.3 Acoustic stealth awareness

Section 2.2.8 introduced ASA as a component of operational effectiveness for military personnel. A substantial gap in knowledge was identified. Specifically, the role of hearing in tasks requiring acoustic stealth is unknown. Hearing might be important in ASA, and it is possible that hearing loss could negatively affect performance during tasks that require acoustic stealth.

Due to the paucity of literature on this topic, this section starts with a description of the terminology that will be adopted. Acoustic stealth behaviour in wider ecology is then considered, which leads to the proposal of how ASA might function in humans. The factors that affect acoustic stealth will then be discussed, which culminates in the identification of broad ASA research objectives.

2.3.1 Terminology

Acoustic stealth refers to the general principle of avoiding detection by audition whilst carrying out an action. In ecology, acoustic stealth is relevant for predators hunting “eared” prey (Goerlitz et al., 2010, Wignall et al., 2011). In warfare, the acoustic stealth of warships and aircraft is considered in order to reduce the range at which a vessel or aircraft can be detected (Wen et al., 2009, McGillvray, 1994).

Acoustic stealth awareness refers to a person’s awareness of their own acoustic stealth. It is distinguished here from *acoustic stealth* by referring specifically to the subjective judgement of one’s own detectability relative to a ‘detector’ (Do I *think* the enemy (detector) could hear me talking?), not the objective detectability (Is the acoustic signature of a soldier’s voice aurally detectability by an enemy?).

Aural detectability refers to the likelihood of detection by audition (Garinther et al., 1985). For example, if a soldier walks on a gravel path 10 m away from an enemy, if the enemy is able to hear the acoustic signature of the soldier’s footsteps then the soldier’s movement is aurally detectable. Aural detectability is distinguished here from acoustic detectability in order to specify detection by the sensation of hearing (aural) rather than detection by microphone (acoustic).

Acoustic stealth behaviour refers to the general behaviours employed by animals to reduce their aural detectability. For example, a lynx hunting a rabbit moves as quietly as possible in order to reach within pouncing distance of the rabbit (Murray et al., 1995); Whilst observing groups of chimpanzees in Uganda, Wrangham et al. (2007) found that groups of chimpanzees reduced their hoots whilst travelling through areas with high risk of detection by rival groups of chimpanzees; Soldiers can adapt their walking style in order to reduce the intensity of the acoustic signature (e.g. the 'Kitten Crawl'[3]).

Aural detectability judgement refers to an individual's judgement of their aural detectability with respect to a nearby enemy. For example, a soldier approaching an enemy judges whether or not the enemy could hear them.

2.3.2 Evolutionary and ecological perspectives on acoustic stealth behaviour

Acoustic stealth is relevant in all aspects of ecology where predator-prey encounters occur. A hunting predator will invariably produce an acoustic signature (e.g. moving through foliage) which can potentially alert the prey to their presence. This puts selection pressure on the prey; the ability to detect a predator by audition is an evolutionary advantage as it might help the prey survive. One example of this is the development of ultrasonic hearing in moths and nocturnal butterflies, which allows them to detect the echolocation used by predating bats and escape (Yack and Fullard, 2000, Rydell et al., 1997). The adaptation of prey to more effective predator-detection abilities requires predators to adapt their behaviour in order to hunt effectively and access food resources. For example, in the case of bats and moths, the *barbastelle* bat emits a reduced ultrasonic echolocation call amplitude that allows them to get significantly closer to moths before being detected (Goerlitz et al., 2010).

A variety of evidence shows that animals make behavioural adaptations to accommodate the acoustics of their actions and their environments. Many species, such as birds, mammals and amphibians, make adjustments to their vocalisations (e.g. spectral changes) in response to background noise in order to maintain effective communications (Egnor et al., 2007, Roy et al., 2011, Bermúdez-Cuamatzin et al., 2009, Boncoraglio and Saino, 2007). In humans, the well-documented 'Lombard effect' describes the tendency of a talker to adapt their vocal intensity to compensate for background noise in order to preserve intelligibility (Zollinger and Brumm, 2011).

The stalking of prey by predators (e.g. a leopard quietly and slowly approaching a deer) typifies what we here define as acoustic stealth behaviour, and is illustrative of the situation that a soldier faces when stealthily approaching an enemy. With no specific study of acoustic stealth behaviour in humans, or wider ecology, to date, it is not clear precisely

---

[3] The 'Kitten Crawl' is a recommended movement technique covered in the 'Basic Battle Skills' pamphlet, which was issued as standard to all infantrymen during the 1950s – 1970s. The pamphlet instructed troops on basic tactics, map reading, first aid and other aspects of infantry service. The Kitten Crawl is described: "It is quiet but slow. It is tiring. Lie on your front, search ahead for twigs" (page 13).

how such acoustic stealth behaviour is acquired. It seems reasonable, though, to suggest that the acoustic stealth behaviour adopted whilst hunting is representative of other animal behaviours in terms of its acquisition. It is therefore worth examining how animal behaviours are acquired as it might give insight into ASA in humans.

Seebacher et al. (2010) argued that animal behaviour is a composite of innate and learned components. They showed this in the hunting behaviour of fish by first giving groups of predatory fish training periods with live prey fish, freshly killed fish or pellets, then observing their predatory behaviour in an experimental arena. They found that all groups of fish attacked with similar frequency, but those with experience of live prey were more effective, suggesting both an innate (all fish attacked) and a learned (more effective with experience) component of behaviour acquisition.

The innate component of animal behaviour usually represents basic stimuli and responses, for example, the acoustic startle response (Davis et al., 1982). Given the complexity of behaviour involved in acoustic stealth, it is likely that this component of hunting is acquired by learning. Various psychological theories of learning, such as habituation, classical conditioning (e.g. Pavlovian conditioning) and operant conditioning, explain how animals interact with and manage the great complexity of stimuli that are presented to them continually in order to achieve favourable, and avoid negative, outcomes (Carlson et al., 2000). In a review of learning in evolutionary psychology, Barrett (2016) suggests that the classic theories of learning cannot explain how an animal might acquire hunting behaviours, such as acoustic stealth, due to the danger associated with learning from one's own experiences. Instead, they suggest that animals possess a social-learning system, which means they can learn the behaviour by observing others perform the task and then imitating. Learning from observing parents and playing with siblings is seen throughout the animal kingdom (e.g. lion cub watching lioness hunt; lion cubs practicing stalking and pouncing on each other). From an evolutionary perspective this mechanism of learning is advantageous, as the cost of failure on tasks requiring acoustic stealth (e.g. stalking prey) could be much greater than learning from observation of an experienced hunter (e.g. their parent). From repeated exposure to the environment, stimuli, and task success within the safer learning environment, the animal can indirectly learn the behaviours required for the real task.

2.3.3 Is effective acoustic stealth behaviour possible in humans?

Most humans living in modern societies need not exhibit acoustic stealth behaviour due to an abundant and easily accessible food supply. The exception is military personnel, whose job might require them to operate with acoustic stealth. Despite this lack of widespread acoustic stealth behaviour, it is reasonable to expect that humans will be able to operate effectively with ASA due to their advanced cognitive abilities and motor control (i.e. they have sufficient control over their bodies to control their acoustic output).

Humans undergo complex cognitive development which enables tasks of far greater complexity than for other animals to be achieved (e.g. communication with complex language, future planning (Carlson et al., 2000)). The enhanced cognitive abilities of some primates and humans relative to other species (e.g. bats, lions) allows us to rapidly adapt our behaviour in order to successfully solve problems and cope with environments and stimuli that we have not encountered before. One classic example of this was the ability of chimpanzees to use their environment to solve a problem. Köhler and Winter (1927) showed this by hanging bananas just out of reach of a chimpanzee in a cage. There was also a box in the cage. After one unsuccessful swipe at the bananas, the chimpanzee then placed the box underneath, climbed on top and reached the bananas. Kohler believed this behaviour showed evidence of insight as the behaviour was different to the trial-and-error approach of cats in similar tasks.

Novel behaviour is able to emerge in humans due to stimulus equivalence (Carlson et al., 2000). This describes the ability of humans to carry out new behaviour that has not been learned or reinforced, based on previous learning. An example of this is teaching a child that a picture of a dog represents the same entity as the spoken word ‘dog’; the child can then also be taught that the written word ‘dog’ is the equivalent entity to the spoken word ‘dog’. The child is then able to learn, by stimulus equivalence, that the written word ‘dog’ is the same entity as the picture of the dog, without being explicitly taught this.

An aspect of cognition in humans (and chimpanzees) that might be relevant to acoustic stealth behaviour is Theory of Mind (ToM) (Premack and Woodruff, 1978). This cognitive ability allows one agent (a person) to infer the mental states (thoughts, feeling and desires) of other agents. This ability to ‘put yourself in another’s shoes’ might be involved for ASA, where the agent (soldier) infers if they are detectable by another agent (enemy) and then acts accordingly to what they believe the enemy might do (e.g. inferring the mental state and thoughts of another). However, given the stalking behaviour of predators that are not known to have ToM (e.g. lions), it is also possible that ToM is not involved. The hypothesised way in which humans have ASA will now be explored, taking into account the various learning and cognitive concepts described here.

2.3.4 Situation awareness during acoustic stealth operations

Situation awareness is a highly important aspect of ground close-combat troops’ operational effectiveness (Endsley et al., 2000, Bevis et al., 2014). Prior to contact with the enemy (e.g. a reconnaissance mission requiring acoustic stealth), the soldier is likely to be engaged in concurrent tasks, such as scanning the environment for enemies using visual and auditory modalities, walking and communicating. Using perceptions of various stimuli in the environment, the soldier is able to continually update their comprehension of the situation and project what is most likely to happen in the short and long term future (Endsley, 1995).

In operations requiring acoustic stealth, situation awareness is critical (Casali et al., 2009). ASA refers to the specific aspect of situation awareness concerned with one's perception and comprehension of their own aural detectability with respective to nearby enemies. We hypothesise that in order for a soldier to maintain situation awareness in acoustic stealth situations (i.e. project if you have been detected by an enemy and are about to get shot), the soldier must incorporate *aural detectability judgements* into their comprehension of the situation. This is a judgement about whether or not the enemy would have been able to aurally detect their presence. We now present a framework for these aural detectability judgements.

2.3.5 Aural detectability judgements

Consider the following fictional acoustic stealth example, based on the task used by Casali et al. (2009). In jungle terrain there is typically ambient noise comprising the sounds of insects, possibly noise generated from wind rushing through leaves and maybe the sound of rain falling on the canopy. A team of soldiers is traversing the terrain during a reconnaissance mission, where they must observe an enemy camp, gather information and return to safety, whilst remaining undetected. Whilst approaching the enemy location they must make ongoing judgements about their detectability to ensure they remain undetected. This will include remaining out of sight (e.g. staying low and wearing camouflage), and remaining sufficiently quiet. They reduce their acoustic signature as much as possible by avoiding noise-making events such as stepping on twigs, talking loudly and loading weapons. Whilst moving they are continually assessing if the sounds they are making (e.g. footsteps, equipment rustling) are likely to be detectable. One soldier steps on a fallen branch, creating a crunching noise that is distinctly different from the sound of the background noise. The soldier (and the other soldiers) must quickly judge if an enemy is likely to now be alert to their position. A highly idealised and schematic version of this aural detectability judgement is shown in Figure 2.1. The soldier stepping on the branch is the subject. The target is the enemy, positioned at some distance away from the target. The subject must make an aural detectability judgement: could the target aurally detect that sound?

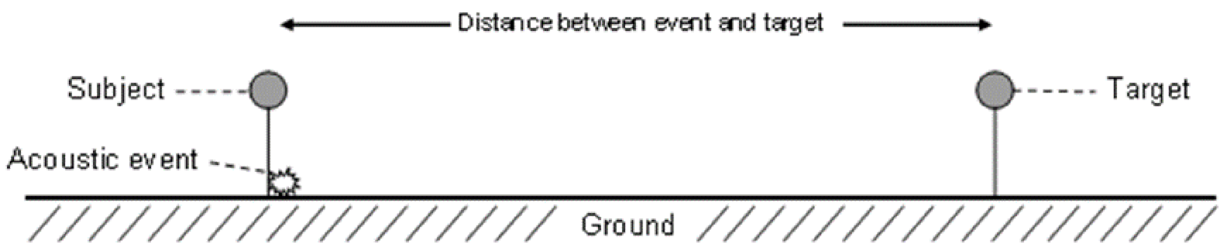

Figure 2.1 - Schematic representation of an aural detectability judgment.

This judgement must be partly based on a cognitive process, not a simple perception-based detection or discrimination task, for two reasons. First, there is a physical limitation which prohibits this judgement from being a simple detection task: the subject only has feedback of the event near the source, not at the target location. Therefore, the judgement must be an inference about the perceptual qualities of the event at the target location. Second, we hypothesise that these aural detectability judgements can take place *a priori* and *post hoc* in relation to the acoustic event. An example of an *a priori* aural detectability judgement is that whilst the soldier walks through the jungle they might see a twig ahead. They know if they step on the twig, a sound might be produced that could be aurally detectable to the enemy; a civilian example includes someone “talking behind someone’s back”, where the talker communicates with someone using suitable vocal effort in order for the speech to be unintelligible or undetectable to the person they are talking about. A *post hoc* aural detectability judgement takes place following the acoustic event, such as when the soldier stood on the branch in the jungle example. The difference between *a priori* and *post hoc* judgements is that the *a priori* judgement is based on a cognitive projection of what might happen, whereas the *post hoc* judgement is based on what has happened.

We hypothesise that an aural detectability judgement (for both *a priori* and *post hoc* judgements) comprises two components. The first is an inference about the physical properties of the event at the target location (incorporating source strength of the stimulus, the stimulus properties, and sound propagation). For *a priori* judgements, this inference is based on prior experience of the potential acoustic event (e.g. what stepping on twigs typically sounds like); for *post hoc* judgements, this inference is based on prior experience (e.g. stimulus familiarity) and sensory feedback of the acoustic event. The second component of the aural detectability judgement is an inference about the perceptual qualities of the event at the target (would the sound at the target location be sufficiently loud for the target to perceive it?).

The justification for this model of aural detectability judgements is based on two factors. Firstly, carrying out this task objectively would first involve calculating the physical properties of the event at the target location by modelling the propagation of the sound, and second determining its detectability by modelling its perceptual qualities (Garinther et al., 1985); therefore aural detectability judgements might incorporate a similar processing stage. Secondly, there is, albeit limited and not directly related, evidence that humans adjust their acoustic output to maintain a constant level at a receiver as the distance between source (talker) and receiver (listener) increases (Johnson et al., 1981, Markel et al., 1972, Zahorik and Kelly, 2007). Zahorik and Kelly (2007) got subject's to say the vowel /a/ for three seconds and adjust the level of their voice such that the level reaching the listener (a microphone) was the same wherever the listener was positioned (1 – 8 m away). This was repeated in a reverberant corridor and an outdoor space. Vocal compensations were found to be within 1.2 dB of the objectively measured propagation losses (compensation needed to overcome loss of intensity by propagation). This was taken as evidence that people have knowledge of how sound attenuates over distance (sound propagation), and that they were able to factor this attenuation into their vocalisations with reasonable accuracy. Given the data suggesting that people might have knowledge of sound propagation, it is possible that people might be able to make reasonably accurate aural detectability judgements, which are assumed to involve an inference based on sound propagation.

2.3.6 Intrinsic and extrinsic factors affecting acoustic stealth

The previous section hypothesised that ASA, at least in humans, involves an aural detectability judgement. There are a multitude of factors that might affect acoustic stealth. Extrinsic factors such as background noise and sound propagation affect the aural detectability of sounds; intrinsic factors such as experience, hearing acuity and attention are hypothesised to affect aural detectability judgements. This section explores these factors.

Figure 2.2 shows a simple schematic diagram of a situation involving acoustic stealth. In this example, assume the subject (a soldier) is approaching the target (the enemy) and is doing so with acoustic stealth behaviour; that is, they are trying to operate without being detected by the target. The star by the subject's foot represents an acoustic event, such as stepping on a twig. The arcs crudely show sound waves, with decreasing weight (amplitude) as they travel farther away from the subject.

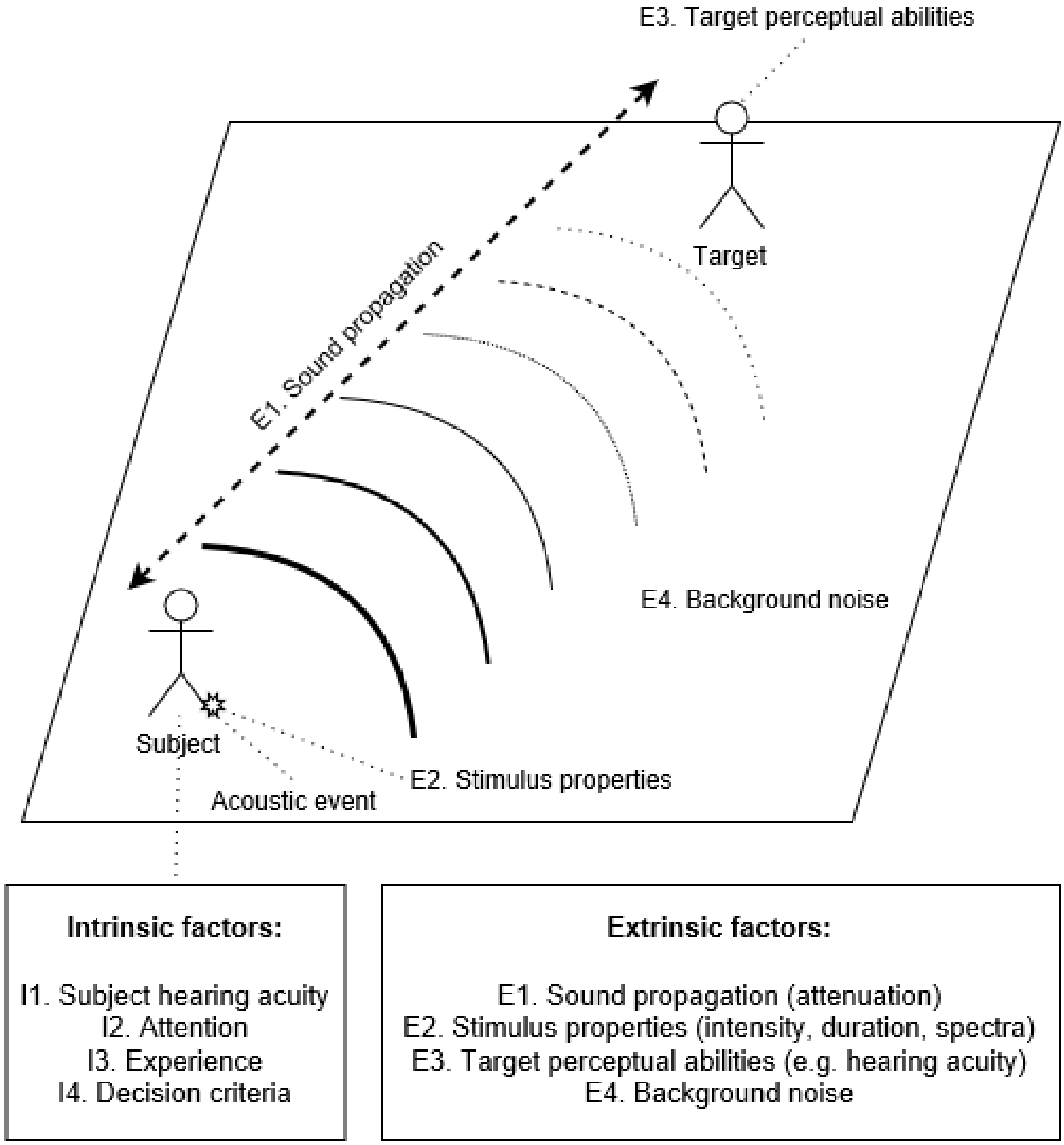

Figure 2.2 - Intrinsic and extrinsic factors affecting acoustic stealth.

2.3.6.1 *Extrinsic factors*

Extrinsic factors are those external to, and out of the control of, the subject and include physical factors (sound propagation, background noise, acoustic event properties) and a psychophysical factor (target perceptual abilities). Sound propagation describes the physical effects of a sound wave travelling from the subject to the target. For a detailed review of sound propagation, see Garinther et al. (1985). As sound is propagated from a source, the sound wave loses energy (attenuation) due to various factors (Wiener and Keast, 1959):

- *Geometric divergence (inverse square law)*: The dominant source of attenuation during sound propagation is the geometric divergence of the sound wave; this accounts for a loss of 6 dB for every doubling of distance from the source (most commonly known as the inverse square law). Therefore, for a source with a fixed sound power, the greater the distance between source and target, the less aurally detectable that source is.

- *Atmospheric absorption:* Sound energy is absorbed by the atmosphere as it propagates. High frequencies are attenuated more than low, and the amount of attenuation depends on the temperature and humidity of the atmosphere, and the distance the sound travels (Bass et al., 1995).
- *Ground effect:* The ground can attenuate the sound during propagation; depending on the distance between source and receiver and the type of ground, the reflections off the ground can interfere with the direct sound path and cause attenuation.
- *Acoustic barriers:* Barriers, such as foliage, trees, walls and buildings also attenuate the sound during propagation.
- *Reflections:* Sound does not attenuate as rapidly with distance in reverberant conditions as in to the free-field (Zahorik and Kelly, 2007).

The acoustic properties (stimulus properties) will affect the aural detectability of the sound by the target. The main factor is the intensity of the sound (source power), as this will affect how far away the sound will be audible. The spectro-temporal properties of the sound will also affect the aural detectability due to the frequency-dependent attenuation during sound propagation and the frequency-dependent hearing sensitivity of the target.

The target's perceptual abilities can also affect acoustic stealth. For example, if the target is wearing hearing protection or has a hearing loss, then the subject will be able to get much closer prior to detection (Price et al., 1989). Fatigue might also affect the target's ability to detect an approaching subject by affecting their vigilance (Matthews et al., 2014).

The ambient noise levels at the target location will have a direct effect on acoustic stealth, as the aural detectability will be affected by masking. In fact, it has been shown that a species of insect exploits environmental noise by attacking spiders more frequently during noisy periods (Wignall et al., 2011).

2.3.6.2 *Intrinsic factors*

Intrinsic factors are those internal to the subject. As previously discussed, we hypothesise that aural detectability judgements comprise two stages: an inference about the physical properties of the stimulus at the target location and an inference about the target's perception of that stimulus. Central to the first stages of a *post hoc* aural detectability judgement is the sensory feedback of the acoustic event, without which the subject cannot judge the aural detectability. For events where the stimulus is created by the subject (e.g. whispering, stepping on a pine cone), there will be sensory feedback in various modalities. For example, if the subject whispers to a fellow soldier, the soldier might get feedback via the tactile and proprioceptive systems as well as auditory feedback; if a soldier steps on a thick branch they will feel the snap through their boot. When the source is not related to the subject, for instance, another soldier nearby shouts to them, the feedback is auditory only. The sensory feedback they receive is critical for estimating the physical properties of the stimulus at the target location, as the level of the stimulus at the source is critical for determining the level at the target.

If the subject has reduced hearing acuity (sensitivity), then they will receive less auditory feedback of the acoustic event. This could potentially lead them to making incorrect aural detectability judgements for two reasons. First, if the sound level is below their absolute threshold for that sound, they will be unaware of the sound and therefore cannot make an aural detectability judgement. Second, their loudness perception of the event will be affected, which subsequently may cause an underestimation of the source strength.

Although purely speculative, it is possible that a hearing-impaired subject might adapt their acoustic stealth behaviour by adopting compensation strategies, such as those observed in the naval command study discussed previously (increased reliance on visual cues, exercise more caution). If true, these compensation strategies will likely come with a cost, such as reduced situation awareness.

On the assumption that aural detectability judgement uses cognitive resources, there might be situations where other stimuli or tasks limit the available cognitive resources to the detriment of ASA. According to Kahneman (1973), we have finite capacity of cognitive resources that can be allocated. Another task which requires allocation of cognitive resources might coincide with the moment that an aural detectability judgement is required; this could reduce the ability of the subject to make an aural detectability judgement. An example of this is that the soldier receives a radio message with grid references that they must remember; whilst listening to the message the soldier steps on a branch but is unaware due to focussing on the radio message.

The role of experience in aural detectability judgements is likely multifaceted. For acoustic stealth behaviour in general (including aural detectability judgements), like most animal behaviour, increased experience will make that behaviour more effective. Therefore, it is reasonable to expect soldiers to have more effective acoustic stealth behaviour than civilians, due to increased experience. For aural detectability judgements though, previous experience of the stimulus is likely to be important. We know that stimulus familiarity is important for auditory distance perception (Kolarik et al., 2016, Zahorik et al., 2005). For example, blindfolded listeners are likely to report that shouted speech is farther away than whispered speech, despite similar presentation levels (Philbeck and Mershon, 2002). Stimulus familiarity should therefore provide an important cue for aural detectability judgements, as prior experience of that stimulus, and in particular how that sound ‘sounds’ at different distances, should contribute to the inference of the stimulus properties at the target location. In this way, one might expect more accurate aural detectability judgements for familiar stimuli than for unfamiliar stimuli. Stimulus familiarity is likely to be critical for *a priori* judgements, as they must be based purely on prior experience of the potential acoustic event.

Aural detectability judgements are hypothesised to be based on inferences. These judgements are therefore being made under uncertainty (e.g. subjects cannot hear what the stimulus sounds like at the target, and they do not know the perceptual capabilities of the target). Subjects making judgements are therefore likely to employ heuristics, which are subject to biases (Tversky and Kahneman, 1974). Heuristics can be thought of as ‘rules-of-thumb’, guiding principles that make complex tasks cognitively simpler. An example of a heuristics in visual depth perception is the *bigger-closer* retinal-size heuristic, where a larger retinal image leads subjects to report that a target is closer (Daum and Hecht, 2009). In aural detectability judgements where the subject has created the sound and has access to sensory feedback of the sound (*post hoc* judgement), a possible is a *louder-further* rule, where louder sounds will travel further and therefore be more aurally detectable. This assumes subjects have knowledge of sound propagation, which a variety of evidence showing level as a primary cue for egocentric auditory distance estimation suggests that they do to some extent (Kolarik et al., 2016). This heuristic may be subject to bias due to individual errors, whereby the subject does not account for sound propagation correctly (rates of attention different to truth), as seen in studies showing greater than 6 dB level changes needed to change egocentric distance by a factor of 2 (inverse square law).

Aural detectability judgments in real acoustic stealth situations are associated with risk, as an incorrect judgement could have costly consequences. Therefore, an important component of the aural detectability judgement is the criterion used in the decision process. For example, if a person was asked to make aural detectability judgements and was told that the consequence of an incorrect judgement (the target could hear the sound but the subject judged the sound inaudible to the target) was certain death, they would be more likely to judge the sound as detectable than if an incorrect judgement had no negative consequence.

In summary, aural detectability judgements have been proposed as an integral part of ASA, where a subject (soldier) makes a judgement about whether or not a target (enemy) would be able to aurally detect the acoustic event. Aural detectability judgements can occur *a priori* and *post hoc* relative to an acoustic event. An aural detectability judgement is hypothesised to involve an inference about the physical properties of the sound at the target location and an inference about the target's perception of that sound. The extrinsic and intrinsic factors affecting acoustic stealth and aural detectability judgements were discussed.

2.3.7 General research aims and the contribution of this thesis

Section 2.2 discussed the role of AFFD in the military and highlighted that very little is known about how soldiers behave in acoustic stealth situations. It was assumed that hearing is likely to be important factor in a soldier's ASA, and so there are fitness-for-duty concerns should a soldier possess reduced hearing capabilities. It was therefore determined that research is required in order to better understand acoustic stealth, and ASA. The previous section proposed that ASA involves aural detectability judgements. Research is now required in order to explore these judgements and acoustic stealth behaviour in general.

Owing to the vast knowledge gap, this section describes the general research aims for ASA. Achieving all of these aims in a single project is unfeasible, but a framework of research is useful to contextualise how the research in this thesis fits into a broader research effort. The proposed general research aims for ASA are as follows.

1. *Develop methodology for measuring acoustic stealth awareness.*

In order to better understand ASA, appropriate methodology needs to be developed. Once a methodology is developed it can be used to explore how the extrinsic and intrinsic factors affect ASA.

2. *Explore extrinsic and intrinsic factors associated with acoustic stealth and the accuracy of a soldier's acoustic stealth awareness.*

There are many factors that might affect ASA, as described in Section 2.3.6. In order to better understand ASA, these factors should be explored using the developed methodology. Of primary interest to the military is whether or not a soldier's ASA is accurate or not (does a soldier know when they're being too loud?).

3. *Characterise the effect of hearing impairment on acoustic stealth behaviour in military personnel and determine the fitness-for-duty implications.*

The main goal of the overall research effort is to understand the effects of hearing loss on ASA, so that the implications for AFFD can be correctly characterised. If it is shown that hearing loss does affect ASA to a level which compromises the safety and operational effectiveness of the soldier (and other soldiers), then ASA should be considered as part of the AFFD testing protocol.

This thesis intends to contribute towards this set of research objectives by developing a method for investigating ASA and using it to explore some of the factors outlined in Section 2.3.6. The following section discusses the proposed experimental approach.

2.4 Experimental approach to acoustic stealth awareness

2.4.1 Overview

Section 2.3 discussed ASA and proposed that aural detectability judgements form an important part of acoustic stealth behaviour. The section concluded with a proposed set of general research aims in order to reach the main goal of understanding the effect of hearing loss on ASA and the implications for AFFD. This thesis intends to contribute to this set of general research aims by developing methodology and using it to explore factors affecting ASA.

The experimental approach and task is first considered, followed by a discussed of the experimental factors that will be explored. A set of objectives required to achieve the proposed method and experimental design is then stated.

2.4.2 Experimental approach

The previous sections in this chapter identified factors that could affect the aural detectability of a sound created by a subject (extrinsic factors) and the factors that might affect aural detectability judgements (intrinsic factors). Several experimental designs were considered in the early stages of this project to investigate ASA, each with their advantages and disadvantages. The dominant challenge was finding a suitable balance between laboratory control and ecological validity. One could re-create an experimental task that requires ASA in an outdoor environment and monitor the acoustic output and behaviour of the soldier, whilst manipulating various factors. This approach has high ecological validity as it accurately recreates the situation and task, but it comes at the high cost of limited control of the acoustic stimuli. The subject is in control of the acoustic stimuli, so there would be difficulty analysing how the stimulus specifically affects the behaviour.

An alternative approach involves recreating an acoustic stealth situation within the laboratory, which permits total control of the acoustic stimuli. However, given the size of available laboratory rooms and the range of distances that are relevant to ASA, it would not be possible to re-create a situation with sufficient ecological validity.

An early idea was to create an experimental task that required subjects to show acoustic stealth behaviour and manipulate the auditory feedback they received. Plans were drawn up to create a circular path of gravel within the University’s large anechoic chamber. The proposed task for the subject was to complete several loops of the path within a certain time, with varying levels of auditory feedback (e.g. with and without ear plugs). Their acoustic output would be measured to see if people are noisier with reduced auditory feedback. This idea was eventually rejected for two reasons. First, because it did not allow us to independently manipulate the auditory stimuli (i.e. the subject was in control). Second, it did not require the subject to make aural detectability judgements with respect to a distant enemy, so the task had poor ecological validity. The requirement of a larger testing environment became apparent, as the subject should ideally make judgements about their detectability by a distant enemy.

Given the requirement of ecological validity, experimental control and a large testing environment, our attention was drawn to technological solutions. An evolving technology that is being increasingly used to simulate various situations in research and training is virtual reality (Parsons, 2015, Bohil et al., 2011, Tarr and Warren, 2002, Kühnapfel et al., 2000). This involves presenting visual stimuli in order to give the impression that the user is in a virtual environment. This technology has been rapidly developing as screen resolutions and the processing power of computers advances, allowing more immersive and realistic experiences to be had. The use of a head-mounted display (HMD), such as an Oculus Rift or HTC Vive, is common in virtual reality systems. With increasing popularity and consumer use, the cost of this technology has decreased hugely.

By using VR presented via a HMD, it would be possible to present visual stimuli to the experimental participant that immerse them in an environment of our choosing (e.g. a jungle, the desert), but to maintain control of the acoustic stimuli as in normal laboratory experiments.

A potential limitation of the use of VR systems is the ecological validity of the experiment. According to the discussion in the previous section regarding aural detectability judgements, the distance between the subject and the target will be a factor that affects acoustic stealth and ASA. It is therefore important that distance perception in virtual reality is similar to that for reality, in order to maintain ecological validity. After conducting a literature review, it became apparent that distance perception over the range of distances of interest (beyond 20 m) had not been tested. Therefore, it was not possible to say with confidence if the use of VR would be appropriate. Given the potential advantages of using VR for ASA research, it was decided that an experiment would be conducted (Experiment 1, reported in Chapter 3). The aim of that experiment was to compare distance estimation in VR and reality in order to validate the use of VR for ASA research.

2.4.3 Experimental task

The experimental task that could be used to explore ASA was considered. As proposed in Section 2.3, ASA is hypothesised to involve an aural detectability judgement, where the subject makes an inference about the physical properties of the sound at the target location and an inference about the target's perception of that sound. It was proposed that the aural detectability judgement could take place *a priori* or *post hoc* in relation to the acoustic event.

It was decided that the experimental task would focus on *post hoc* aural detectability judgements. This was chosen because it is most critical for soldiers to make correct decisions when their acoustic stealth has actually been compromised (i.e. there was an acoustic event that would be aurally detectable). Measuring an individual's aural detectability judgements also lends itself to a relatively simple experimental task: present a stimulus to a subject and obtain their judgement about the detectability of that sound to a target. The experiment can then be used to explore the effect of various factors on the judgements. Of primary interest to the military is the accuracy of a soldier's acoustic awareness. In terms of aural detectability judgements, this means comparing the subject's aural detectability judgement to the true aural detectability. The experimental task should therefore be based on an aural detectability judgement.

2.4.4 Psychophysical approach

2.4.4.1 *Expectations*

It is expected that there is a relationship between the level experienced by the subject and the subject's aural detectability judgement, which is characterised by a psychometric function (PF). Suppose the subject's task was to view a target in the distance, listen to a sound and judge whether the target be able to detect that sound, responding yes or no. A hypothesised PF showing the relationship between the level of the sound and the

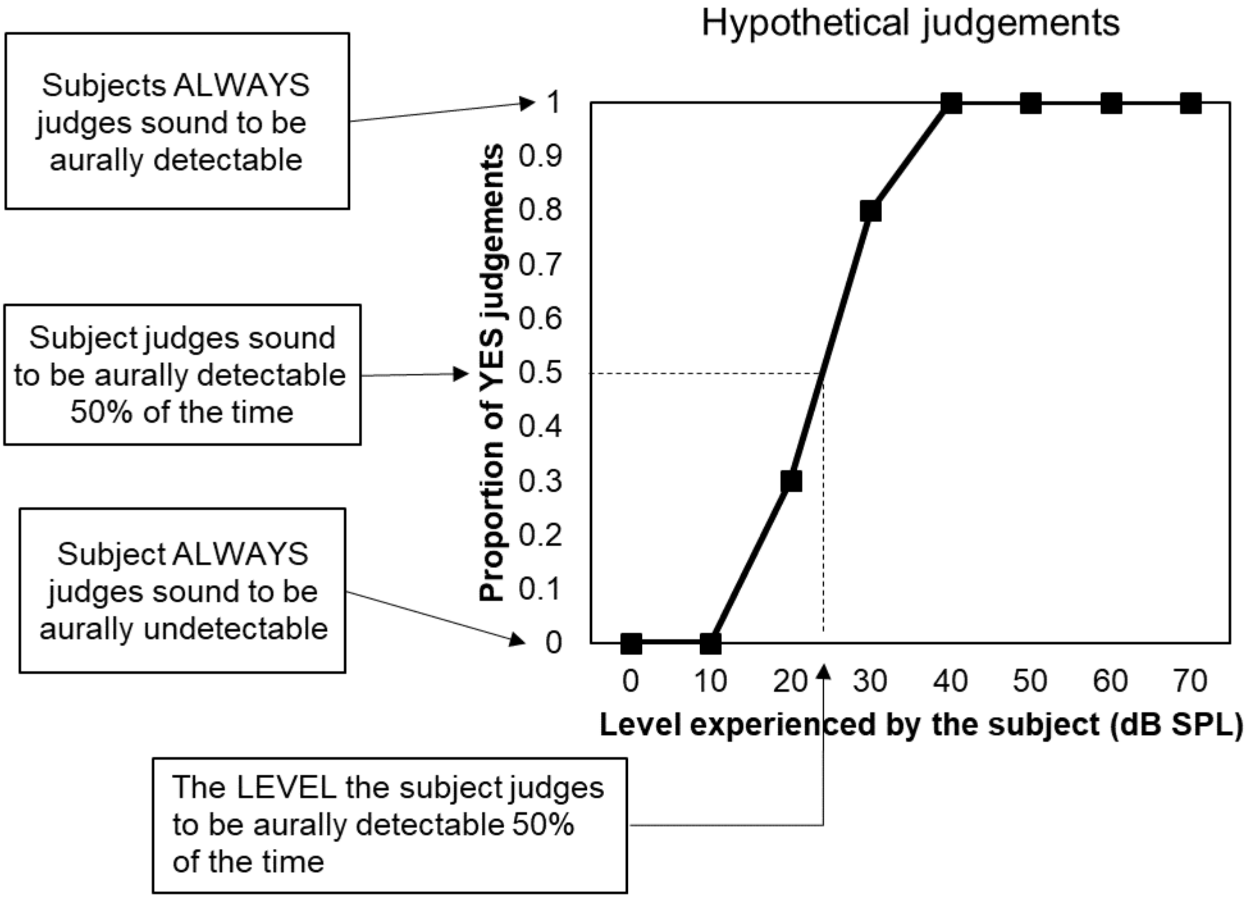

Figure 2.3 - Hypothetical relationship between stimulus magnitude and aural detectability judgements.

proportion of times they respond ‘Yes’, p(yes), is shown in Figure 2.3At high levels, it is expected that the subject will always respond yes. This is because, presumably, very high levels, such as those created by a weapon (e.g. grenade), will result in the subject responding ‘Yes’ all the time (p(yes) = 1). Therefore, the upper asymptote of the PF will reach 1, assuming no lapses in concentration on the task.

At levels close to or below the subject’s threshold for that sound, it is expected that the subject will respond ‘no’ all the time (p(yes) = 0). This is because they should factor their perception of the sound into their judgement of the target’s perception of the sound. If the sound is so quiet that they cannot, or they can only just, detect it, they would judge the sound to be undetectable to the target. If this expectation is true, the subject’s hearing thresholds will modulate their judgements, such that a subject with a hearing loss (loss of sensitivity) would be unable to judge sounds below their threshold as they themselves have not detected the sound. In terms of the psychometric function, it is expected that a hearing loss would move the function towards higher stimulus levels.

The nature of the task, a single interval yes/no task, means that the lower asymptote of the PF should be at p(yes) = 0, providing the subject was presented with a sound at a sufficiently low level.

The mid-point of the PF reflects the stimulus level that the subject judges 50% of the time that the target would be able to detect the stimulus. Although a somewhat arbitrary definition, this could be taken as the subject's judged level that corresponds to the target's threshold level.

2.4.4.2 *Method selection*

Most classic psychoacoustic methods would permit measurement of aural detectability judgements, such as method of constants, adjustment, limits and adaptive procedures (Prins and Kingdom, 2016); however, each has different advantages and disadvantages. The method of adjustment would involve allowing the subject to vary the level of the sound until they think the target would just be able to detect the sound. This would allow a rapid assessment of the location of the PF, but it comes at the cost of no information about the slope of PF, and the location estimation may not be as accurate as for the method of constants (Gescheider, 2013). The method of limits and adaptive procedures also have the advantage that the stimulus level that corresponds to the target's threshold (location of PF) could be quite rapidly estimated; however, the information about the slope of PF using these methods is limited.

During the early stages of determining which method to use, a simple pilot experiment was designed that used the method of adjustment. This took place in an approximately 50-m long corridor within a university building. The subject sat on a chair at one end of the corridor with a loudspeaker placed directly behind them. They viewed a target in the distance (a manikin on a stool), who was stood at various distances. The subject was presented with bursts of white noise, 1 second in duration, with a gap of 1 second between each burst. The subject used the scroll wheel of a computer mouse to adjust the level of the noise (1 dB change per scroll increment). Their task was to adjust the noise to the level they judged to be equivalent to the quietest level the target (in the distance) would be able to detect. The final level that they selected was taken as the judged level equivalent to the target's threshold. This was measured twice at various distances (5, 10, 20, 40 m). Following the subject's judgements, the procedure was repeated, but this time the subject viewed the author placed in the same locations as the manikin target, who gestured via hand signals whether they could detect the noise; this found the 'true' aural detectability level at each target distance (the author had normal hearing). The subject was then given feedback about their judgements. Following feedback, the subject repeated their judgements, with two judgements for each distance. Figure 2.4 shows the set-up.

Figure 2.4 - Subject completing pilot experiment.

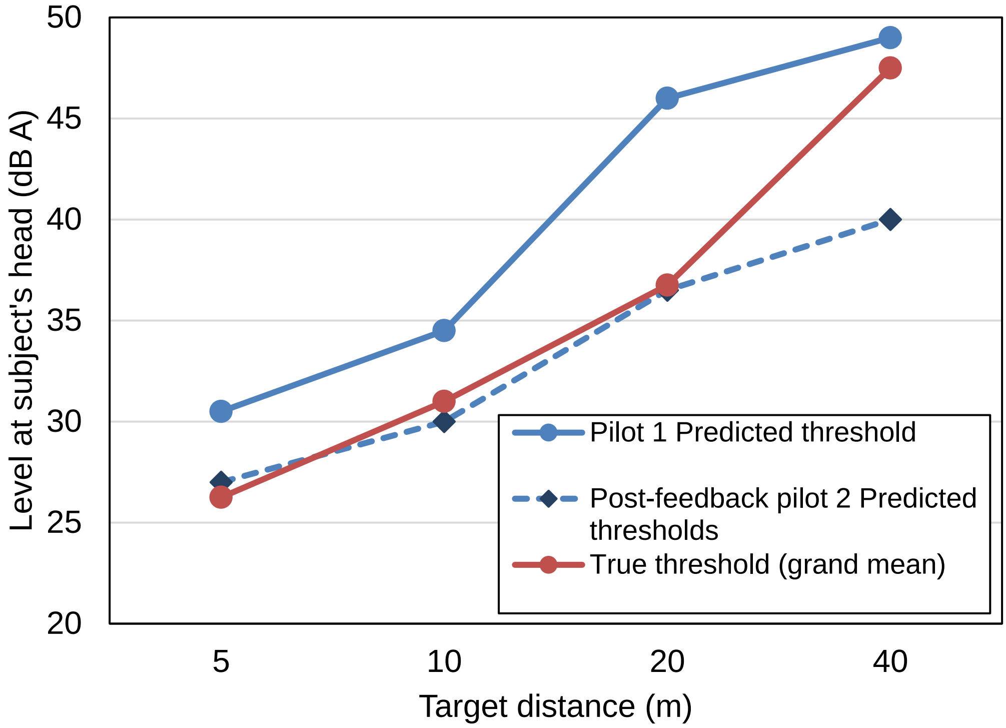

Figure 2.5 - Pilot experiment results (n = 1).

Figure 2.5 shows the subject's judgements and the 'correct' target thresholds. The results showed that the subject factored the target distance into their judgements, with increasing level as the target was further away. The accuracy of judgements for the first set of judgements (Pilot 1 predicted thresholds) was poor, with the subject judging that the target would not be able to detect the sound, when they were able to. Following feedback, the judgements became much more accurate. The 'true' thresholds show a greater change between 20 and 40 m, than between 10 and 20 m and 5 and 10 m. This is most likely because of increased background noise at the far end of the corridor, which raised the target actual threshold. There was an average of 5 dB difference between repeats within one condition (target distance).

Although from just one subject and with several flaws in the experimental design, these data suggest that aural detectability judgements are sensitive to stimulus level and target distance. It is also interesting that judgements were not accurate on the first judgement and improved with feedback.

The task and method (method of adjustment) were readily accepted by the subject, who reported understanding the task with ease and found the method of giving judgements straightforward. However, limited information is available about judgements at levels other than the final adjusted level. Thus, the method of adjustment is not ideal for learning about aural detectability judgments, and ASA. It also does not have ecological validity, as a soldier's task during an operation is not to adjust their acoustic output to a level they judge to be just audible.

Based on the insights this pilot study gave and the information that this method missed, it was decided that the method of constant stimuli would be used to measure aural detectability judgements. Given that this study is the first of its kind, it was considered sensible to use a method that captures information about the PF. The probable non-uniform effect of background noise on the judgements gives further motivation to using VR, as total control of the visual and acoustic environment can be achieved.

2.4.5 Experimental factors

2.4.5.1 *Stimulus level*

Given that the aural detectability of a sound depends on the source power (Garinther et al., 1985), the stimulus level experienced by the subject gives an indirect cue to the true aural detectability of the sound. Subjects' judgements should therefore be sensitive to stimulus level. If the results of the experiment show that subject's judgements are sensitive to stimulus level, it also shows, to some extent, that the experiment had construct validity (the ability of the test to measure what it purports to measure). The psychoacoustic approach (method of constants) allows this factor to be explored, and therefore the effect of stimulus level on aural detectability judgements was explored.

2.4.5.2 *Target distance*

The distance between subject and target is a major factor in determining the aural detectability of a sound. Given the expected importance, the target distance (the distance between source and target) was explored in the experiment. It was predicted that subjects would incorporate sound propagation into their judgements, such that they would judge a particular stimulus level (as experienced by the subject) to be less detectable as target distance increases, e.g. for a stimulus level of 40 dB SPL at the subject location, they would judge the stimulus to be less detectable by a target at a far distance than a close target, because the sound 'fades' as it travels (in layman's terms).

The range of distances that were relevant to acoustic stealth were determined in a number of ways. During the first year of the PhD project, the author visited the Brecon Infantry Battle School, which is a major Army training centre. During the visit, the author observed several exercises involving a squadron attacking an enemy, both with blank and live ammunition. One particular exercise required the soldiers to attack an enemy camp in the middle of the night. The ambient noise was very low; there was no wind and the location was far away from any sources of traffic noise. The soundscape was highly similar to the 'stillness' of an anechoic chamber. This involved stealthily approaching the known enemy location and attacking. The distance that the troops started their attack from was approximately 50 – 75 m (i.e. the distance from the target that they used acoustic stealth behaviour to reach). This gave an approximate range of target distances of interest for experimental work. However, the target distances are limited by the ecological validity of the VR environment. Therefore, the target distances that were tested depended on the results of Experiment 1.

2.4.5.3 *Stimulus type*

As previously discussed, the accuracy of ASA is of primary interest to the military. It was therefore considered important that the types of sounds used in the experiment were ecologically valid in order to assess accuracy for sounds they might encounter whilst in a real situation. Also, given that we propose that aural detectability judgements rely on knowledge of how sound propagates, real-world (familiar) stimuli should improve the accuracy of the judgements. In order to explore if the familiarity of the stimulus does affect aural detectability judgements, a synthetic non-familiar stimulus (stationary white noise) was included.

2.4.5.4 *Test-retest reliability*

Given the novelty of the task and the method, it was considered important that a measure of test-retest reliability was included, for two reasons. Firstly, for future researchers, the precision of the measurement (aural detectability judgments) is useful to know for sample size calculations. Secondly, very poor test-retest reliability gives information about either the judgements or the measurement that would not be known unless it was measured.

An ideal test is one that gives identical scores when the same person is tested on different occasions under identical conditions (Summerfield et al., 1994). In practice this is rarely the case due to measurement error. The error associated with the measurement arises from systematic (learning, fatigue) and random (errors that summate to randomly affect the test result, e.g. spontaneous firing of auditory nerve, attention of subject) effects (Bland, 2015).

In order to investigate test-retest reliability, the same subject must repeat the measurement on multiple occasions under (as close to as is possible) identical conditions. Then, three analyses can be performed to quantify the test-retest reliability. Firstly, a repeated-measures ANOVA can be performed, with repeat as a source of variance; a significant effect of repeat shows that there was a systematic difference between the repeats. Secondly, the correlation between pairs of measurements shows the replicability of the scores (Summerfield et al., 1994). Finally, the repeatability, “the value below which the difference between two measurements will lie with probability 0.95” (Bland, 2015), is identified by calculating the standard deviation of the differences between scores measured on separate occasions under identical conditions, multiplying by 1.96 and dividing by the square root of 2. This calculation assumes that the variability is homoscedastic, each measurement is independent of each other and the error associated with each measurement is equal.

2.4.5.5 *Hearing impairment*

As stated in the general research objectives, understanding how hearing impairment affects ASA and an individual’s AFFD is the primary motivation for the research. Due to the novelty of the proposed experiment, it was considered too risky to recruit hearing-impaired military personnel for the first study of its kind. The use of hearing-loss simulation to overcome recruitment difficulties was considered. However, the findings of Experiment S3 (Section A.4) suggested that the available real-time hearing loss simulation device would not be appropriate. The ‘offline’ versions (Experiments S1 and S2, Sections A.2 and 0) would potentially be feasible in a virtual acoustics experimental rig; this idea could be explored in future studies, but not in this project.

2.4.5.6 *Military experience*

As discussed previously, experience of a behaviour generally increases the effectiveness of the behaviour. Therefore, it would be expected that military experience would improve the accuracy of judgements. The inclusion of military personnel was considered in depth and in light of the general research objectives associated with ASA research. Due to the novelty of the method and the difficulty in recruitment of military participants, it was decided that this experiment would focus on civilian personnel in order to make progress on the development of the methodology, which is the first general research objective. Depending on the results of this project, testing military personnel should be a priority of future work.

2.4.5.7 *Additional factors*

There are factors additional to those discussed above that are of acute interest to the military. These factors include: tinnitus, which might have a distracting effect which compromises a soldier’s ability to make aural detectability judgements (Yankaskas, 2013); and hearing protection devices, which will reduce the auditory feedback available for aural detectability judgement (Casali et al., 2009). These factors are beyond the scope of the current work, so will not be discussed in detail.

2.4.6 Summary of experimental approach

This section discussed the experimental approach that was adopted in order to explore ASA. The following main points were discussed:

- In order to explore ASA, situations in which factors affecting acoustic stealth should be simulated, so that accuracy of ASA, and the various factors proposed that might affect the judgements, can be explored.
- The use of virtual reality was proposed as a means of presenting visual stimuli independently of auditory stimuli, thereby allowing laboratory control whilst maintaining ecological validity; however, distance perception in VR has not been tested over the range of target distances of interest (> 20 m), so an experiment is required to explore this and validate the use of VR.
- The experimental task is based on *post hoc* aural detectability judgements, where the subject listens to a sound and makes a judgement of the aural detectability of that sound by a target.
- The psychophysical method was the method of constant stimuli as this should provide information about the expected psychometric function.
- The experimental factors that were explored include:

- The accuracy of judgements
- Stimulus level
- Target distance (distance between subject and target)
- Stimulus type
- Test-retest reliability

2.5 Summary and aims

Hearing is a crucial component of an individual's fitness for duty in the military. As such, AFFD measures are used to identify individuals considered unfit for duty. We argue that the current AFFD measures used in the UK Armed Forces lack predictive validity and that improve measures are required.

Acoustic stealth awareness is an important component of operational effectiveness for those who encounter ground close-combat. Soldiers can find themselves in situations where they must remain undetected by a nearby enemy. ASA refers to the soldier's awareness of their aural detectability, that is, the likelihood they will be detected by the enemy. We proposed that ASA is achieved using aural detectability judgements; that is, a subject makes an inference about the physical properties of the sound at the target location (e.g. at the enemy's ears) and the target's perception of that sound (was it detectable?). The accuracy of these judgements is important, as appropriate action is required if a soldier's acoustic stealth is compromised.

A significant knowledge gap for ASA was identified, and three general research objectives were identified. First, a methodology for investigating ASA should be developed. Second, the accuracy of, and factors affecting, aural detectability judgements should be explored. Finally, the effect of hearing impairment and the implications for AFFD should be investigated. Achieving these objectives requires a program of research, and this project aims to contribute towards these objectives. Specifically, this project aims to develop a method for investigating ASA and explore how factors affect judgements and their accuracy.

The experimental approach was discussed and it was decided that the method would measure aural detectability judgements. The experiment task involved playing a sound to the subject and making them judge whether that sound would be detectable by the target. The method of constant stimuli was used to explore various effects on judgements, including: stimulus level, target distance and stimulus type. The test-retest reliability was measured. The experiment involved a virtual reality system in order to simulate a visual environment much larger than that available in a laboratory. A validation experiment was required to assess whether this was appropriate.

The aims of this thesis were therefore:

1. Design suitable methodology for measuring aural detectability judgements
2. Explore the effects of various factors on aural detectability judgements
3. Evaluate the accuracy of aural detectability judgements

Chapter 3 Visual distance estimation in virtual reality and reality

3.1 Experiment 1: Introduction

Recall from Section 2.4 that virtual reality (VR) headsets allow independent control of visual and auditory stimuli in an experiment, and allow large and outdoor environments to be simulated within a laboratory setting. These qualities make VR a potentially powerful tool for acoustic stealth awareness (ASA) research, where large environments are required[4]. Recall also that it is hypothesised that aural detectability judgements are modulated by the distance between subject (solider) and target (enemy) in an acoustic stealth situation (Section 2.3.6). Therefore, if VR is to be used for ASA experiments, distance perception must be similar in VR and reality for the results of the experiment to be valid.

Most studies of distance perception in virtual reality have focussed on egocentric[5] distance estimates in personal and action space (Renner et al., 2013). The perceptual space surrounding a person can be divided into three circular, egocentric regions (Cutting and Vishton, 1995): *personal space* (up to 2 m), *action space* (up to ~30 m) and *vista space* (beyond ~30 m). Distance estimates typically show depth compression in virtual reality compared to estimates in reality, of around 10 – 20%, meaning that subjects tend to underestimate distance (Armbrüster et al., 2008, Knapp and Loomis, 2004, Witmer and Kline, 1998). The data constituting the general picture of underestimation are all from distance estimates in action space (up to 20 m). The underlying reason for this compressive effect is still unknown and subject to debate in the literature (Foreman et al., 2004, Buck et al., 2018, Richardson and Waller, 2007). Some research has challenged whether the underestimation is actually due to the VR modality. A study by Knapp and Loomis (2004) replicated the limited field of view inherent in head mounted displays (HMDs) by systematically reducing participants' field of view in a distance estimation task. The results showed that changing the field of view had no effect on the estimates, suggesting that the limited field of view in HMDs cannot explain why underestimation seems to occur in VR. Creem-Regehr et al. (2016) suggested that the inconsistency observed in the underestimation between studies might be explained by the method of

[4] 'Large environments' are used here to refer to environments where a soldier may encounter acoustic stealth situations, such as woodlands, urban areas and jungles.
[5] Egocentric distance is the distance between you and the target. Exocentric distance is the distance between two objects, neither of which are you.

estimation and the environment. Their data showed systematic differences between the methods of distance estimation for both indoor and outdoor environments, and that underestimation was greater for outdoor environments.

In contrast to egocentric distance perception in action space, very little is known about performance in vista space. One possible reason for this is that most VR applications do not depend on faithful distance perception in vista space, such as surgical training (Seymour et al., 2002, Kühnapfel et al., 2000) and neuroscience research and rehabilitation (Tarr and Warren, 2002, Bohil et al., 2011), and so there is no particular motivation to carry out this research.

There is good reason to expect that distance perception in vista space would be similar in VR to that in reality due to the availability of distance perception cues in both viewing environments. For judging the distance of objects close to a person (e.g. the distance between the floor and your eyes) both monocular cues (pictorial cues, motion parallax) and binocular cues (retinal disparity, stereopsis and convergence) are used (Carlson et al., 2000). In contrast, only monocular cues are useful in vista space. Providing the virtual environment accurately preserves the visual cues available in the real version of the scene, it is expected that distance estimation will not exhibit the systematic underestimation effect observed in action space. This expectation is based on a study by Loyola (2018) who showed that increased availability of visual cues led to more accurate distance estimates.

Due to limitations in the screen resolutions of HMDs, it is reasonable to predict that the richness and availability of cues cannot be fully preserved, especially when judging very distant objects. A primary cue for judging distant objects is retinal image size, since an object’s physical image size on the retina decreases as the object moves further away (Gilinsky, 1951). In VR HMDs, for an object of a given size, there will be a threshold egocentric distance where the screen is unable to preserve the changes in retinal size, due to limitations in screen resolution within the HMD. Therefore, it is expected that beyond a threshold distance, depending on the size of the object being viewed, there will be a divergence between estimations made in VR (using HMDs) and reality. This expected divergence has not yet been explored in the literature. It might also be the case that the screen limitations generally reduce the richness of cues (e.g. texture gradient is less distinguishable in VR than in reality). Therefore, although the same visual cues are present, a subject might have to work harder to make estimates.

Several methods can be employed to measure distance estimates (for an extensive review of methods, see Knapp (2003)). These include verbally reported absolute estimates where the subject views the target and gives a numerical estimate (in metres or feet) of the distance between themselves and the target (e.g. Daum and Hecht (2009)). Other measures include action-based tasks, such as blind-walking where the subject typically views the target for a period, then walks to the object whilst being blindfolded (e.g. Rieser et al. (1990)) and perceptual matching, where the subject views the target and has to perform a perceptually driven task, such as identifying the mid-point. For distance estimation in vista space, verbal report is the most commonly used method due to the practicalities of measuring the distances involved.

As with distance estimation in virtual reality, distance perception in vista space in real environments has received significantly less scientific investigation relative to personal and action space. Most studies have been carried out in a laboratory using a projector, which limits the ecological validity of the results (Fine and Kobrick, 1983, Bee, 1991). Only one study has explored distance estimation in vista space (> 20 m) in an outdoor environment. Daum and Hecht (2009) found large between-subject variability in verbally reported estimates for various target distances: for a target distance of 54 m, the mean estimate from 12 subjects was 54.3m with a standard deviation of 25.3 m. Similar variability was observed in all conditions and experiments reported in the study, suggesting that distance estimates in vista space are highly variable. In the first experiment they showed that subjects tended to overestimate the distance by 25 – 100 % over target distances of 54 – 460 m, with increasing relative error as the target distance increased. In another experiment they found that subjects underestimated when the target distance was less than 75 m by up to 20%, and overestimated for target distances greater than 75 m, by up to 60%. The results showed that accuracy of judgements is highly variable between subjects and, on average, they tend to underestimate distances less than 75 m and overestimate distance beyond 75 m.

Since no data are available to determine whether distance estimation is similar in VR to reality in vista space, this study aimed to contribute to this gap in knowledge. Due to the effects of method and environment on distance estimates (Creem-Regehr et al., 2016), the primary aim of the study was to assess whether distance perception in VR is *similar* to that for reality, using the same method for both modalities and a similar environment. For example, if someone tends to underestimate distance by 20% in reality, do they also underestimate distance by 20% in virtual reality? The absolute accuracy of estimates in the real environment is also of interest, in order to evaluate how our method that for compared to previous studies.

It was expected that, due to the ability to preserve relevant visual cues to distance perception, estimates in VR should be similar to estimates in reality. Providing the inter-pupillary distance is calibrated for each individual user, VR HMDs are able to preserve binocular effects (stereopsis, binocular disparity) to a large extent (Loyola, 2018). VR

HMDs are also able to preserve most monocular cues, depending on the extent to which they are present in the environment. In order to maximise the chances of similar distance estimates between environments, it was considered important that extensive effort should be taken in order to preserve the cues in VR that are available in reality. These cues included: texture gradient, motion parallax, relative size, target familiarity and adequate lighting.

This rest of this chapter describes a study of egocentric distance estimation in VR and reality in vista space. It was predicted that distance estimates would be similar in VR to reality, due to the availability of relevant cues in both environments. A self-report measure of the workload associated with the task was included, as the resolution limitations of the screen in the VR headset may reduce the richness of cues, thereby increasing the effort required to perform the task. It was predicted that the self-reported effort associated with the task would be higher when estimating distance in VR than in reality.

3.2 Methods

3.2.1 Participants

Young healthy adults were recruited to the study. Their ages ranged from 18 – 35 years. All subjects had normal or corrected-to-normal vision and were naïve to the purpose of the study.

3.2.2 Test location

We chose an open grassy field on Southampton Common as the test location, for two reasons. Firstly, we wanted the data to represent generic distance estimates, rather than those specific to a particular location with a unique set of landmarks or cues (e.g. along a path or in a sports hall). Secondly, the location was sufficiently remote to permit undisrupted testing yet situated conveniently close to the University. The field was flat and lined by trees only, and did not have any distinct landmarks, such as lone trees or shrubs.

3.2.3 Stimuli and design

Subjects made distance estimates in two environments: real and virtual. The VR environment was a virtual version of the Southampton Common test location, created using Unity 5.5 (Technologies, 2016) and viewed using an Oculus Rift Consumer Version 2 (CV2) head-mounted display. For full details of the how the environment was designed, see Appendix C. The virtual environment condition of the experiment took place in a university laboratory. The reality environment condition took place on Southampton Common. A screenshot of the virtual environment can be seen in Appendix C.

The experimental task was the same in the two environments. The subject viewed a target and estimated the egocentric distance (distance between themselves and the target), in metres. The target that the subject viewed was a human (the author). The height of the target was 1.70 m. Five target distances (25, 50, 75, 100 and 125 m) were used, resulting in a 2 (environment type) X 5 (distance) repeated-measures design.

In both environments, 3 blocks of trials were completed. Each block consisted of 22 trials comprising the 5 target distances at 3 different lateral locations (15 trials), plus 7 additional target distances between 18 and 85m (see Figure 3.1). The lateral locations were at -10°, 0° and 10° horizontal azimuths relative to the subject’s forward direction and were used to reduce effects of landmarks unique to one particular azimuth (e.g. a particular tree might aid estimates on one particular azimuth). The 7 additional target distances (‘filler’ trials) were used to prevent bias or habituation to the 25 m step size. The order of the 22 trials was randomised in each block. The first block in each session was a familiarisation period and was not included in the analysis. The estimates at each distance for blocks 2 and 3 were analysed, resulting in 6 estimates per distance for each environment (2 blocks x 3 azimuths per distance in each block). The mean of the 6 estimates was used in the data.

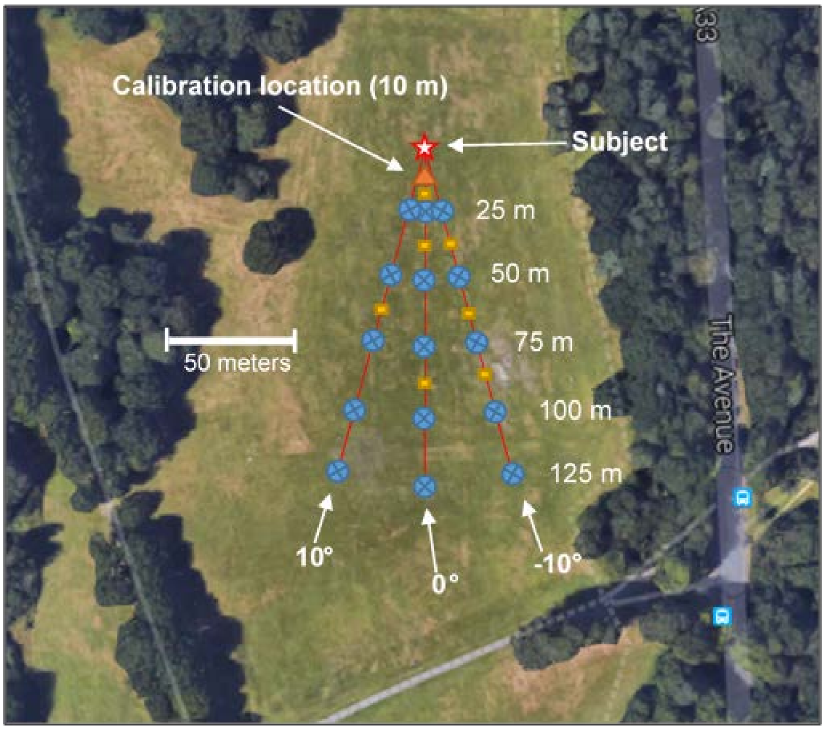

Figure 3.1 Target location schematic. Satellite image of test site annotated with target locations. The subject's location is represented by the Red star. The 15 target locations are shown by blue circles, at distances 25 -125 m. The yellow rectangles represent the 7 additional target locations. The orange triangle shows the calibration location. Image provided by Google Maps.

The target locations were measured using a surveyor's measuring tape. The locations were marked by a small peg that was invisible to the subject.

The experimental design was given ethical approval by the University ethics board (ERGO ID: 26625).

3.2.4 Procedure

The experiment comprised two sessions carried out on separate days. One session was for the VR condition and the other for the reality condition. To remove order effects, half of the subjects completed the virtual environment session first and the other half completed the real environment first. In each session, 3 blocks of trials were completed. To help calibrate the subjects' estimates, the subject viewed the target at 10 m at 0° azimuth before each block. The calibration exercise was identical to the method used by Daum and Hecht (2009). During the calibration exercise, the subject was then instructed to estimate the distance between themselves and the target. The subject was then told the true distance (10 m). The block of trials then proceeded.

During each trial, the subject viewed the target and estimated the egocentric distance, in metres. The subject gave their response by writing their estimate on a notepad. For efficiency, two subjects completed the reality environment condition simultaneously. To prevent the subjects from influencing each other's estimates, they were not allowed to talk during the experiment and wrote each judgement on a new page of the notepad. Subjects could have as long as they needed for each trial; typically, a trial lasted 8 - 15 s. No feedback was given for the experimental trials. Once the subject had given their estimate, an assistant got the subjects to turn around so they could not see the target. The target then moved to the next target location and informed the assistant via radio that they were in position. The assistant then got the subjects to turn around and start the next trial. The virtual environment condition procedure was identical to this except that the subject gave their distance estimate via verbal report, and the HMD screen faded to black in between trials.

After each block of trials, the subject completed the NASA-Task Load Index (NASA-TLX) questionnaire to give a measure of workload during that block of trials (Hart, 2006). The questionnaire consists of six subscales: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. For each subscale, the subject gave a rating on a 100-point scale with 5-point steps, based on the task they had just performed. The 'raw' unweighted ratings are analysed here.

Subjects received £20 for completing the experiment. An additional £20 was given as a prize to the subject with the most accurate estimates to help motivate the subjects.

3.2.5 Statistics

For the primary statistical analysis, the results of the experiment were converted into relative estimation errors (%). These express the estimate as a percentage of the true distance, e.g. an estimate of 110 m for a true distance of 100 m would be a 10% estimation error. The use of relative estimation error is required due to the heteroskedastic data observed in most distance perception studies. Converting the estimates to relative estimation error makes the data homoscedastic. The estimates were analysed using a repeated-measures analysis of variance, with viewing environment (reality and VR) and distance (25, 50, 75, 100, 125 m) as factors.

Based on the standard deviations observed in Daum and Hecht (2009), a power calculation based on Bonferroni corrected paired sample t-tests (number of t-tests = 5, adjusted $\alpha = 0.01$, $\beta = 0.2$) estimated that a sample size of at least 18 subjects was required in order to detect a 10% difference between conditions, assuming the standard deviations are similar in VR to those in reality (Faul et al., 2007).

3.3 Results

3.3.1 Distance estimates

The mean estimates and standard deviations for 21 subjects are shown in Table 3-1. Estimates are expressed in both their absolute form (in metres) and as estimation error relative to the true distance (percentage). Of the 24 subjects that took part in the experiment, 1 subject withdrew from the experiment and 2 subjects were excluded from the analysis as their estimates were extreme outliers. Subjects' estimates from blocks 2 and 3 were pooled to increase the reliability of the measurements. The mean of the 2 estimates was used for all analyses.

The average distance estimates showed near-veridical accuracy in both viewing environments, with a tendency towards a slight underestimation of distance (see Figure 3.2 and Figure 3.3). The average underestimation was most pronounced in the virtual environment for the 125 m target distance (-10.8% estimation error). There was a large amount of between-subject variability, as seen in the large standard deviations; however, the variability was similar between viewing conditions and across distances.

A repeated-measures ANOVA (2 environments x 5 distances) for the mean estimate (average estimate from blocks 2 and 3) was conducted. Mauchly's test showed that the assumption of sphericity had been violated for both distance and the interaction between environment and distance; therefore, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity for the main effect of distance and the interaction. The ANOVA revealed no significant effect of environment, $F(1,20) = 0.102$, $p = 0.753$, effect of distance, $F(1.798,35.952) = 1.454$, $p = 0.247$, or interaction between environment and distance, $F(1.881,37.622) = 2.692$, $p = 0.084$.

The figure and near-significant interaction term of the ANOVA suggested there might be a trend for target distance to affect estimates differently in each viewing environment. This was explored by testing the linear contrast trend across distance for each viewing environment, which showed a significant trend for the virtual environment, $F(1,20) = 8.484$, $p = 0.009$, and no trend for the normal environment, $F(1,20) = 0.101$, $p = 0.754$.

Table 3-1 - Mean distance estimates, expressed as absolute estimates (M) and relative estimation error (%), and standard deviations (SDs).

Distance (m)	**Environment**	**Absolute estimate (m)**			**Relative estimation error (%)**	
		M	***SD***		***M***	***SD***
25	Real	24.8	7.0		-0.8	28.1
	Virtual	26.0	7.8		3.9	31.3
50	Real	48.6	14.4		-2.9	28.8
	Virtual	49.3	14.9		-1.4	29.8
75	Real	72.1	26.2		-3.9	35.0
	Virtual	70.6	23.9		-5.8	31.8
100	Real	96.2	45.1		-3.8	45.1
	Virtual	91.1	33.5		-8.9	33.5
125	Real	120.8	53.1		-3.4	42.5
	Virtual	111.6	43.0		-10.8	34.4

Mean estimates expressed as absolute estimates (m). Solid black line shows estimates in reality; grey dotted line shows estimates in virtual reality. The thin dotted black line shows veridical distance. Error bars show ±1 SE.

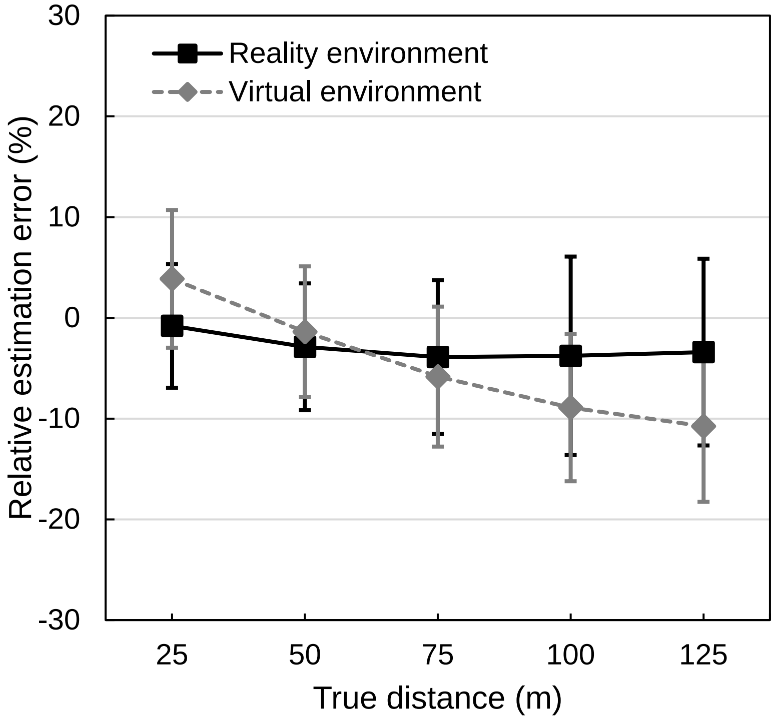

Figure 3.2 - Relative estimation error of distance estimates. Black squares and line show the estimates in reality; Grey diamonds and dotted line show estimates in virtual reality. Error bars ±1 SE

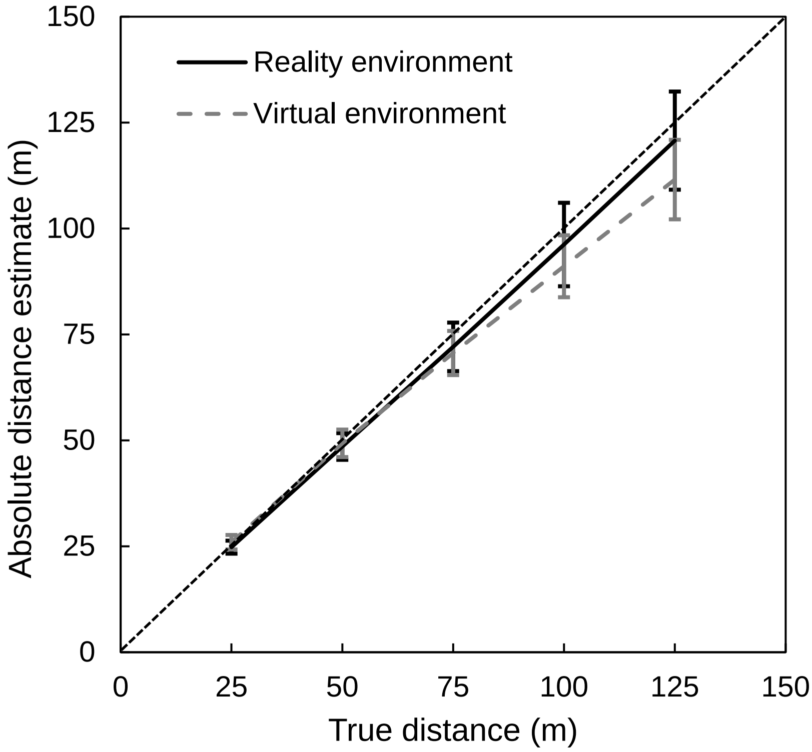

Figure 3.3 - Mean estimates expressed as absolute estimates (m). Solid black line shows estimates in reality; grey dotted line shows estimates in virtual reality. The thin dotted black line shows veridical distance. Error bars show ±1 SE.

3.3.2 Questionnaire results

The NASA-TLX questionnaire results are reported in Table 3-2 and displayed in Figure 3.4. The questionnaire responses from blocks 2 and 3 were pooled. Wilcoxon signed-rank tests were used to compare each domain of the NASA-TLX as the assumption of normality was violated. None of the domains of the NASA-TLX revealed significant differences between viewing environments, suggesting that the task-induced workload was similar in the two environments.

Table 3-2 - NASA-TLX questionnaire results. The median and interquartile range, and the Wilcoxon signed-rank test p-value is shown for each domain.

Domain	**Reality** *Median (IQR)*	**Virtual reality** *Median (IQR)*	**Wilcoxon test** *p*-value
Mental demands	37.5 (22.5 - 45)	37.5 (22.5 - 70)	0.184
Physical demands	10 (2.5 - 22.5)	10 (5 - 15)	0.189
Temporal demands	12.5 (7.5 - 20)	15 (5 - 17.5)	1.000
Performance	57.5 (40 - 60)	55 (40 - 70)	0.323
Effort	37.5 (20 - 50)	50 (17.5 - 70)	0.150
Frustration	12.5 (0 - 32.5)	15 (2.5 - 22.5)	0.950

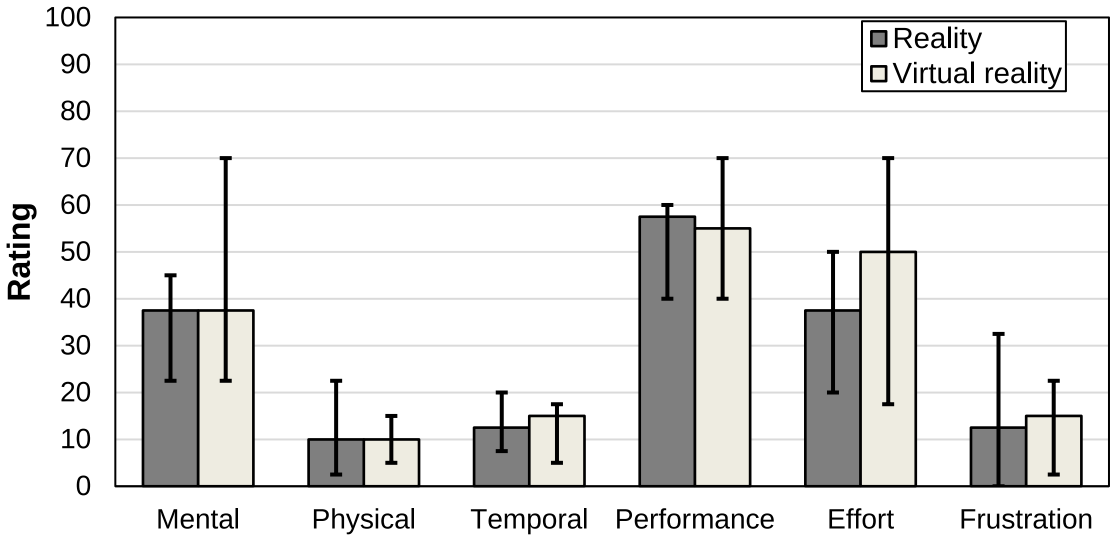

Figure 3.4 - NASA-TLX ratings. Dark grey bars show median ratings for reality environment; light grey bars show median ratings for virtual reality environment. Error bars shown the interquartile range.

3.4 Discussion

3.4.1 Overview

The primary aim of this experiment was to compare distance estimates in vista space in virtual reality and reality. The results showed near veridical average estimates in both environments, with a linear trend for VR where estimation error decreased as target distance increased. However, no statistical difference between the two viewing environments (VR vs real) at any target distance (25, 50, 75, 100 and 125 m) was found, suggesting that there is not a statically robust difference between VR and reality, at least under the conditions tested in this experiment. The findings with respect to caveats and limitations are discussed in greater detail below.

3.4.2 Comparison to other studies

The distance estimates that people gave in the reality conditions of this experiment were, on average, more veridical than found in other studies employing similar methods. As discussed in the introduction to this chapter, Daum and Hecht (2009) (D&H) is the only directly comparable study. They found that near distances (25 – 75 m) were generally underestimated (-10 to 0% error) and far distances (> 75 m) were generally overestimated (0 – 50% error). The results of the current experiment are not grossly different from those for the D&H study, but the differences are worthy of discussion.

The method used in the current study was similar to that of D&H. However, a few key differences may explain the difference across studies. First, both this experiment and the D&H study showed large between-subject variability, shown by the large standard deviations for each estimate in both studies. The variability was heteroskedastic in both studies when estimates are expressed in absolute terms (in metres), whereby the magnitude of the standard deviations increased as the target decreased. The D&H study only required subjects to make one estimate per target distance, whereas the current experiment involved six estimates per distance (plus three practice estimates). In the current experiment we used the mean of the six estimates to represent each participant's estimate. Using the mean of six estimates would reduce the within-subject variability, meaning the predominant source of variability in the dataset was the between-subject variability. In the D&H study, both within- and between-subject variability may have been large, so their results may not be as representative of a subject's true distance perception. As D&H obtained only one estimate per condition, it is not possible to analyse their data in order to compare within-subject variability, so it is difficult to conclude with certainty whether or not this explains the differences between studies.

Another difference between methods that might explain the inconsistent results is the target type. D&H used white wooden boards with black crosses, whereas we used a human. As a result, subjects in the current experiment could use the retinal image size visual cue more effectively because of target familiarity (Fitzpatrick et al., 1982, O'leary and Wallach, 1980).

Given the large between-subject variability, the sample size is of particular importance; The sample mean could deviate quite substantially from the population mean with a small sample size. The D&H study used sample sizes of 12 and 15 in their field studies, whereas we used 21 subjects. It is possible that the D&H sample means estimate the population mean less accurately due to the smaller sample size, although our sample size was not much larger.

3.4.3 Similar estimates in virtual reality to reality

The statistical analysis of distance estimates collected in this experiment suggested that the estimates did not differ for VR and reality. There was also no effect of distance, or interaction between distance and viewing condition, on the relative estimation error, suggesting that the accuracy of the estimates was similar at each distance. However, there is a limit to what can be concluded from these data due to the power of the experiment. A *post hoc* power calculation, using the observed means and standard deviations in each viewing condition showed that with 80% power this experiment would have been able to detect a difference of 10.5% in estimation error (using a Bonferroni corrected paired-samples two-way t test). Therefore, it can be concluded that, using the environment and method utilised in this experiment, it is unlikely that there exists a difference in the average distance estimation errors between VR and reality of greater than 10.5%.

The generalisation of these results to environments substantially different to this, such as densely populated scenes (jungles, forests), or highly sparse scenes (deserts, open ocean) is not possible. However, for open space environments that preserve monocular cues, such as texture gradient, retinal size, relative size and target familiarity, and for the methods used here, the results here should generalise well.

3.4.4 Task load scores

The NASA-TLX revealed no differences in any domain measured by the questionnaire. The hypothesis that reduced richness of cues as a result of hardware limitations in the VR headset would result in greater effort during the task cannot be accepted, based on these data. These results are corroborated by the anecdotal and informal comments from subjects during the experiment that suggested they found both tasks easy and non-effortful.

3.4.5 A trend towards divergence?

The results showed a linear contrast trend for decreasing estimation error as target distance increased for the virtual environment, but not the normal environment. These results are limited to the range of target distances measured here (25 – 125 m), meaning it is unknown whether the trend continues beyond the range of distances estimated here. Although the ANOVA revealed no significant difference between viewing environment, there may be a divergence between viewing environments (see Figure 3.3 and Figure 3.2). As discussed in the introduction to this chapter, this might be related to screen resolution limitations. Consider the height of a human at various distances, and how that relates to the angular size of the light from the human that reaches a viewer’s retina. As the target human moves further away from the viewer, the angular size (or retinal size) of the target human reaching the viewer’s retina decreases in an inversely proportional relationship. At far distances (e.g. 500 m), the angular size changes minutely when the target human changes from, say, 500 to 550 m. This is a problem with HMDs, as the screen resolution limits the relative changes in size that can be achieved. Therefore, the ability of VR to preserve cues at far distances will differ from reality. Although not statistically robust, the trend for a divergence may be explained by limitations in HMD resolution. This is sceptical, however, and requires further work, especially if VR is to be used in situations involving identifying a human-sized target beyond 125 m.

3.4.6 Implications for research using virtual reality

The primary aim of this experiment was to compare egocentric distance estimates in VR and reality in vista space. The findings show that for applications using an outdoor open environment with distance perception cues similar to those in a real environment, distance estimation is unlikely to be different, on average, by more than 10.5%, compared to an equivalent real environment. The findings are generalizable over distances of 25 – 125 m and for human targets, and will only be appropriate if effort is made to preserve available cues. As previously discussed, these results do not generalise to environments with substantially different cues, such as densely populated scenes like jungles. Additional cues(e.g. occlusion) in these environments may improve the accuracy of judgements, but research would be required in order to assess that.

3.4.7 Implications for acoustic stealth awareness research

A primary motivation for carrying out this experiment was to determine whether the use of VR HMDs is appropriate for ASA research. The results suggest that distance estimation in VR is likely to be within 10.5% (estimation error) of reality, using the equipment, environment, and methods selected, and over the target distances used here. Therefore, the use of VR for ASA research is indicated providing the following caveats are considered:

1. The virtual environment and target used should not differ greatly from that used in this experiment.
2. The similarity in distance estimation was indicated on average, but not necessarily on an individual level, so results of the ASA experiment should be analysed for average trends and caution should be exercised when considering individual performance.

3.5 Summary and conclusions

This chapter described Experiment 1, which compared egocentric distance estimations in VR and reality in an outdoor environment for target distances between 25 and 125 m. The experiment was conducted because of a lack of evidence comparing distance estimation in VR and reality in vista space (> 30 m), which potentially contraindicated the use of VR for acoustic stealth awareness research. It was also intended to contribute to the identified knowledge gap of VR distance estimation in vista space.

The experiment tested 21 subjects using a verbal report method, and found near veridical accuracy when the average estimates were considered. The estimates had similar between-subject variability to previous studies and the estimates in reality were more veridical than those in other studies, which may be explained by differences in methods. There was a tendency for a divergence between estimates in VR and reality as the target distance increased beyond 75 m, which might be explained by hardware limitations.

In summary, the conclusions of this experiment were:

- Differences of greater than 10.5% are unlikely to exist between average distance estimates in reality and virtual reality, in an open outdoor field environment, when estimating the egocentric distance to a human target stood 25 – 125 m away.
- The use of a virtual reality headset in acoustic stealth awareness research is suitable, providing the same virtual environment, target and target distances used in this experiment are utilised.

Chapter 4 Acoustic stealth awareness: Method development

4.1 Introduction

Recall that in Chapter 1 acoustic stealth awareness (ASA) was hypothesised to involve an aural detectability judgement, shown schematically in Figure 4.1. This requires the subject to make a judgement about whether a target would be able to detect a sound that occurs near the subject. Various methods were considered, and partially developed, culminating in a method proposal outlined in Section 2.4. The factors that might affect ASA, and subsequently aural detectability judgements, were discussed in section 2.3.6. The factors that this experiment will focus on are the distance between subject and target and the stimulus type. This chapter describes the methodological developments that took place in order to measure aural detectability judgements (Experiment 2, reported in Chapter 5) and achieve the aims set out in Section 2.4.6.

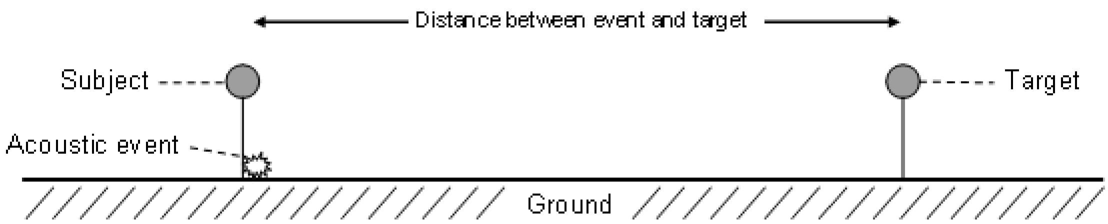

Figure 4.1 - Schematic representation of aural detectability judgement.

This chapter discusses the following:

- The psychoacoustic method for in experiment 2.
- The development of stimuli for experiment 2, including the selection, production, recording, processing and equalisation of stimuli.
- The development of the hardware and software to be used in experiment 2.
- Pilot studies of the methodology

4.2 Psychoacoustic method

In section 2.4.4, it was decided, following a pilot experiment, that the psychoacoustic method used in the experiment would be the method of constant stimuli. Within a given condition (target distance, stimulus type), this involves presenting the subject with a sound at various levels repeatedly and measuring the proportion of 'yes' responses at each level. The number of levels to test, the number of repeats, and at what level to present were all unknown. These were each determined by pilot experiments. It was decided that the initial number of levels would be 6. The number of repeats at each level that would be piloted was 8. The levels presented are discussed in the report of the pilot studies in section 4.5.2.

4.3 Stimuli

Stimuli were generated according to the process shown in Figure 4.2. The following sections describe each stage of the process.

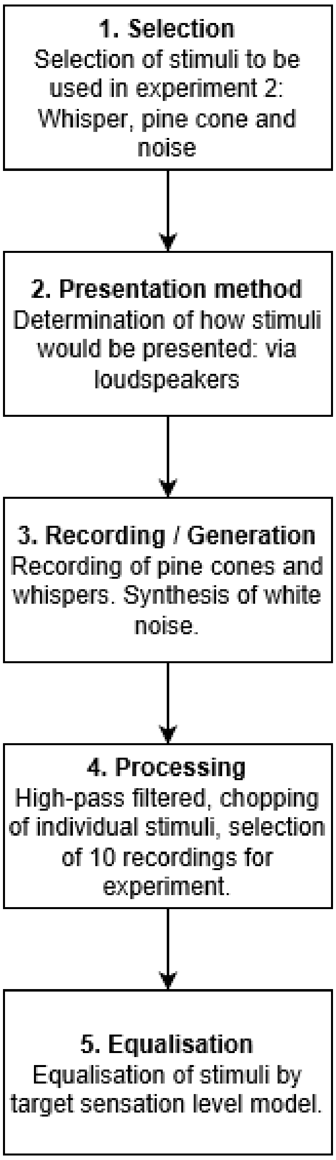

Figure 4.2 - Process for the generation of stimuli used in experiment 2.

4.3.1 Stimulus selection

The aim of the experiment was to explore how basic factors affect aural detectability judgements and to assess the accuracy of judgements. One factor of interest was the type of stimulus, as the judgements may be dependent on the type of stimulus, given the known effects of stimulus familiarity of egocentric auditory distance judgements (Section 2.3.6.2) and how previous experience of a sound could aid aural detectability judgements (Section 2.4.5.3). To address the accuracy research question whilst maintaining ecological validity, it was considered important that the type of sounds used were representative of those experienced in situations requiring ASA. During the previously described visit to Brecon Beacons Infantry Battle School (Section 2.4.5.2), various ecologically valid sounds were found to be associated with acoustic stealth situations. The sounds generally fit into three categories:

- Interactions with the environment
- Communications
- Interactions with equipment

In acoustic stealth situations, soldiers typically have to interact with their environment to achieve the goal of the operation. This might include locomotion, which produces sound when their feet touch the ground. Their equipment may also produce sound, such as the ‘rustling’ of their clothing whilst moving. For a team of soldiers to operate effectively, some communication typically occurs. This might involve silent hand gestures, but at times can involve whispering or radio communications, as was observed during the Infantry Battle School visit. Finally, a soldier sometimes must interact with their equipment, such as loading their weapon or opening and closing pockets.

Based on the observed sounds, it was decided that a sound related to an interaction with the environment and a communication sound should be used as the types of sound in the experiment. The two stimuli selected were the sound of stepping on a pine cone (hereon referred to as ‘pine cone’) and a whispered digit. A third sound type, stepping on a twig, was selected in addition to the other two; however, pilot testing suggested that there was too much testing time per participant, so the twig was excluded from testing.

An additional type of sound was selected in order to explore the effect of stimulus familiarity. As discussed in section 2.3.6.2, it was hypothesised that prior experience of the sound (familiarity) might aid the accuracy of judgements. It was therefore decided that an unfamiliar sound should be included, to allow a comparison between familiar sounds (whispers/pine cones) and unfamiliar sounds. A synthesized broadband white noise (hereafter referred to as ‘noise’) of similar duration to the whispers was included.

4.3.2 Stimulus presentation

Reproducing these sounds whilst precisely manipulating the level can only be achieved via headphones or loudspeaker. The alternative is to get the participant to create the sounds themselves, but this does not allow the desired control of the stimulus. This is acknowledged as a methodological weakness, as reproducing the sound via a loudspeaker or headphones does change the overall sensation of that sound, because the participant will not get any visual, tactile, or proprioceptive feedback of that sound. These judgements will therefore be based exclusively on the auditory feedback and not multisensory perception.

It was decided that stimuli would be presented via loudspeaker. Firstly, the location of the sound source is important for determining the 'true' detectability of the sound, due to the inverse square law of sound propagation (Garinther et al., 1985). This requires the participant to correctly externalise the sound source in order to get valid auditory feedback of the acoustic event. Although externalising sound sources via headphones is possible, the results can vary between listeners and a long process of measuring individual head-related transfer functions is usually required (Brungart, 1998, Zahorik, 2002, Brimijoin et al., 2013, Wenzel et al., 1993). By using a loudspeaker, the issue of externalisation of the sound source was avoided.

The position of the loudspeaker needed to be ecologically valid. Whispered speech could occur by the subject (the person performing the aural detectability judgement) whispering or a nearby person whispering to the subject. To reproduce the subject whispering using a loudspeaker, the loudspeaker would have to be placed close to the head, similar to loudspeaker placement in some echolocation studies (Papadopoulos et al., 2011, Rowan et al., 2017). This would reduce the ecological validity of the whispered speech, and might become uncomfortable for the subject during long testing sessions. These problems can be overcome by simulating the latter situation, where another person whispers to the subject. The loudspeaker can then be placed at the other person's mouth to accurately simulate the source location. It was decided that to avoid problems related to the directionality of the loudspeaker vs the true source, the loudspeaker would be placed directly behind the subject, with the subject intersecting the direct sound path between source and target. This loudspeaker placement simulates a situation where there is another soldier stood behind the subject and they whisper a number.

A reasonable assumption is that the sound of someone whispering to you and the sound of stepping on a pine cone will rarely have co-located sound sources. Using a pine cone sound in the experiment therefore requires an additional loudspeaker. It was decided that the pine cone loudspeaker should be located close to the foot of the subject to recreate the subject stepping on the pine cone. In order to stop the subject from occluding or touching the loudspeaker with their feet, and to get the loudspeaker close to the true source location, the subject stood on a platform and the loudspeaker was placed underneath. The height of the platform was 0.25 m, so that the loudspeaker could fit underneath. See Section 4.5.3 for further details about the platform.

For the noise, it was decided that the sound would be reproduced from the loudspeaker that produce the whisper. This decision was based on the planned familiarity comparison between the noise and whisper. Note, the height of the loudspeaker producing the whispers was set to 1.95 m, so that the cone of the loudspeaker was in the position that simulates the mouth height of an average male (1.75 m). This height takes in to account the height of the platform. A schematic diagram showing a side-on view of the loudspeaker arrangement is shown in Figure 4.3.

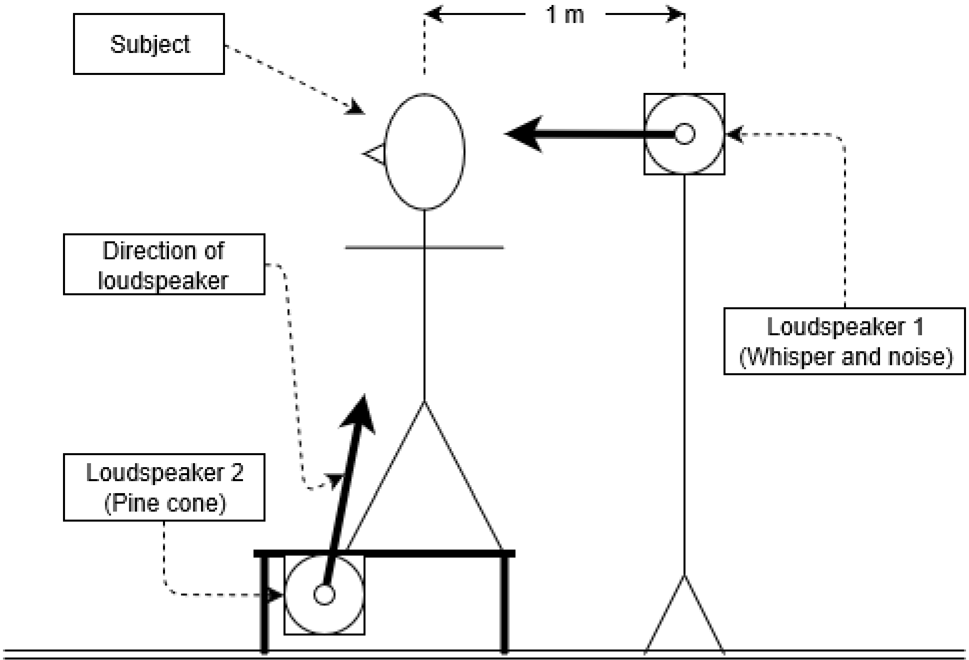

Figure 4.3 - Loudspeaker locations for experiment 2. Large black arrows indicate the direction the loudspeaker is facing. Subject is facing forward, looking in the direction of right to left, and Loudspeaker 2 is 1 metre directly behind.

4.3.3 Stimulus recording

Following the selection of stimuli and presentation methods to be used in experiment 2, the stimuli were now generated or recorded. For the whispers and pine cones, the stimuli were recorded in the small anechoic chamber at the University of Southampton. Pine cones (approximately 100) were collected from the New Forest National Park. The voice used to record the whisper was the author’s.

The small anechoic chamber was set up according to Figure 4.4. A carpet tile was used to simulate soft ground underfoot. This helped recreate the sound of an outdoor pine cone crunching. Each microphone used was a Bruel and Kjaer Type 4193 Falcon. For the whispers, only the whisper microphone was present. The microphone was placed 1 m away from the listener's mouth. No pop filter was used since the distance between the source and microphone rendering it unnecessary. 1 m was used as the source-microphone distance because this would be the distance between source and subject in the experiment.

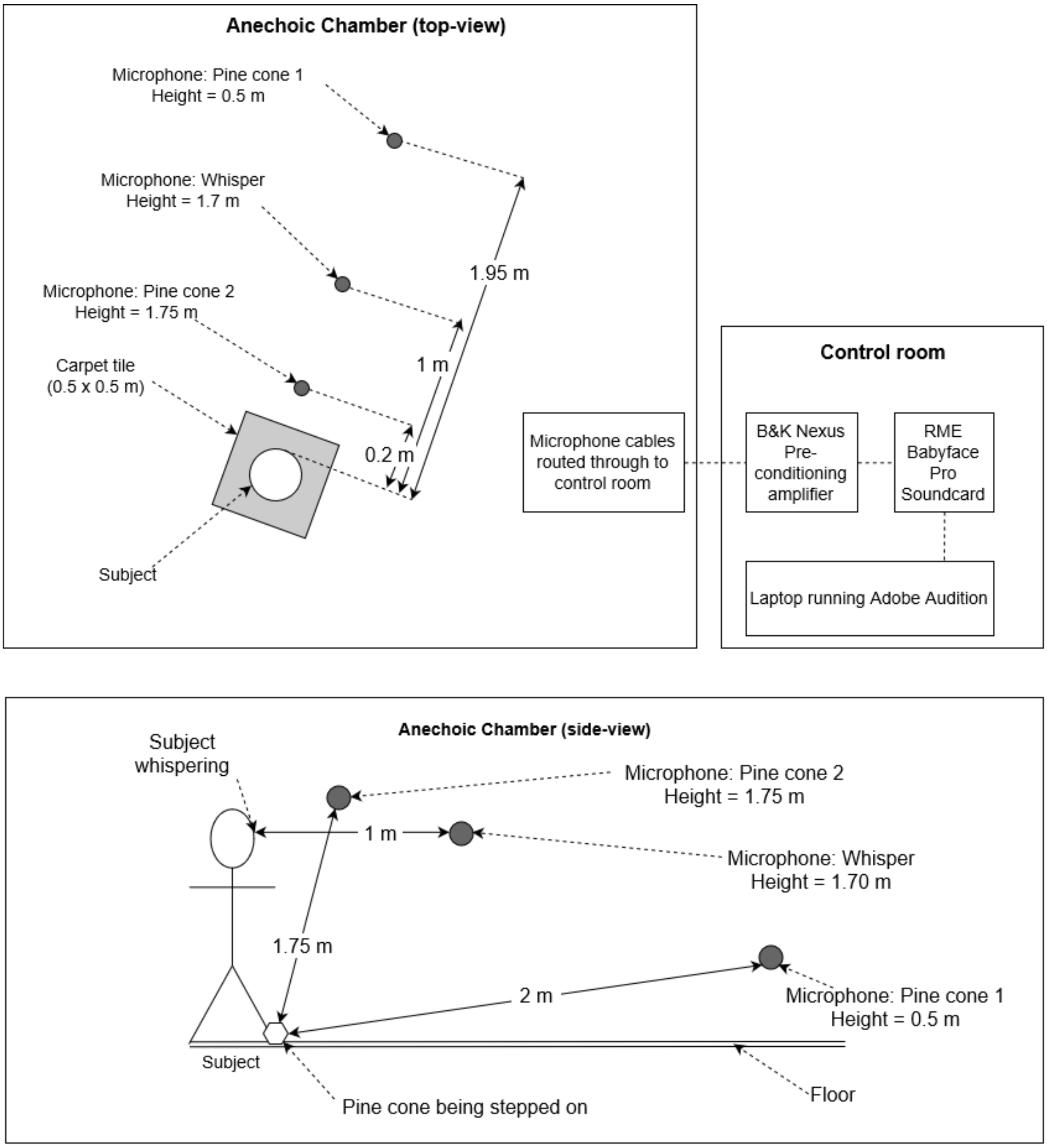

Figure 4.4 - Schematic diagram of recording set up. The top two panels show a top view of the anechoic chamber and the control room. The bottom panel shows a side view of the anechoic chamber. Microphone locations are shown by dark circles. Diagram not to scale.

For the pine cones, the use of two microphones was essential for two reasons. For each pine cone recorded, Pine cone 1 was used to obtain a calibrated measurement so that the sound propagation model could be used in order to predict the level of the sound at various target locations (model described later). The height of Pine cone 1 was 0.5 m in order to better capture any directional effects of the sound source. Pine cone 2 was placed near the subject's head in order to get a calibrated measurement of the level at the subject's head for that same pine cone recording. This meant that for each pine cone, the level at the subject and the level at the target could be calculated.

Each microphone was routed to a Bruel and Kjaer Nexus Preconditioning amplifier, then to a RME Babyface sound card and finally into the laptop, which recorded the audio using Adobe Audition. Pine cones and whispers were recorded in several long audio files (approximately 30 minutes). The gain of the Nexus amplifier was set to maximise the signal-to-noise ratio whilst avoiding clipping. Before each recording, each microphone was calibrated using a Bruel and Kjaer Sound Calibrator (Type 4231), which produced a 1 kHz tone at 94 dB SPL when coupled to the microphone.

For the pine cones, each of the approximately 100 pine cones were stepped on. The author stepped on the pine cones as if walking slowly, in order to replicate stealthy movement. For the whispers, the author spoke each digit (one to ten, excluding seven) approximately 15 – 20 times.

4.3.4 Stimuli processing

Four stages of processing occurred in order to create the stimuli for Experiment 2. Firstly, each recording (approximately 30 minutes) was roughly broken down into individual pine cone or whisper sound files. Secondly, each individual recording was carefully (manually) shortened so that there was between 50 and 100 ms prior to the start and end of the main signal. Thirdly, each sound file was band pass filtered with lower cut off of 20 Hz and higher cut off of 20,000 Hz, and cosine-squared ramps of 45 ms were applied at the start and end. To stop the loudspeakers from clicking, each file was zero padded with 100 ms of zeroes at the start and end of the file.

Finally, each sound file was listened to in order to select the stimuli that would be used during the experiment. As discussed previously, ten repeats at each level would be used in the pilot experiments. If the exact same recording were used for each repeat at each level, the subject might adapt to that particular signal similar to the “frozen noise” effect (Felty et al., 2009). Also, the ecological validity of the aural detectability judgements would be reduced, as the judgements would be based on one example of that stimulus type. A solution was to use a different example of each stimulus type. Providing there are not large differences (spectro-temporal, timbre, etc.) between examples, the potential adaptation and limited ecological validity issues can be alleviated. Ten samples from each stimulus types were selected, based on their subjective quality.

For the pine cones, ten recordings were selected that had similar duration and were subjectively similar in terms of their ‘crunchiness’. For the whispers, ten recordings were selected that were considered to subjectively have similar vocal effort and similar waveforms (selected by the author). The digits “Three” (n = 2), “Four” (n = 3), “Five” (n = 3) and “Nine” (n = 2) were selected, as they subjectively had similar temporal envelopes and timbre; “One”, “Two”, “Six” and “Eight” were excluded because the intensity of the consonants relative to the rest of the word reduced the homogeneity of those digits compared to the others (Three, Four, Five and Nine).

The noise was generated using MATLAB code written by Dr Daniel Rowan. This code generated Gaussian noise of the desired duration, which was then band pass filtered with lower cut off of 20 Hz and higher cut off of 20,000 Hz. Wav files were generated with 45 ms cosine-squared ramps at the start and end of the noise. 100 ms of zeroes (silence) were added to the start and end of the wav file to avoid clicking from the loudspeaker.

An example waveform and its respective spectrogram for each stimulus type is shown in Figure 4.5. The calibrated levels of the recorded stimuli (n = 10 per stimulus type) are shown in Table 4-1. For the whispers, the levels shown in the table represent the level as measured 1 m from the source, as shown in Figure 4.4. For the pine cone, the levels represent the level close to the subject’s head, as shown by Pine Cone 2 in Figure 4.4.

Table 4-1 - Mean, standard deviation, minimum and maximum levels for recorded whispers and pine cones as recorded (n = 10 recordings per stimulus type). All levels are in dB SPL re. 20 µPa.

	Whisper		**Pine cone**	
Statistic	***RMS***	***Peak***	***RMS***	***Peak***
Mean	49.4	65.7	60.0	88.4
Standard deviation	3.8	3.9	2.9	4.4
Minimum	44.1	58.1	56.7	82.1
Maximum	54.5	71.1	65.2	97.3

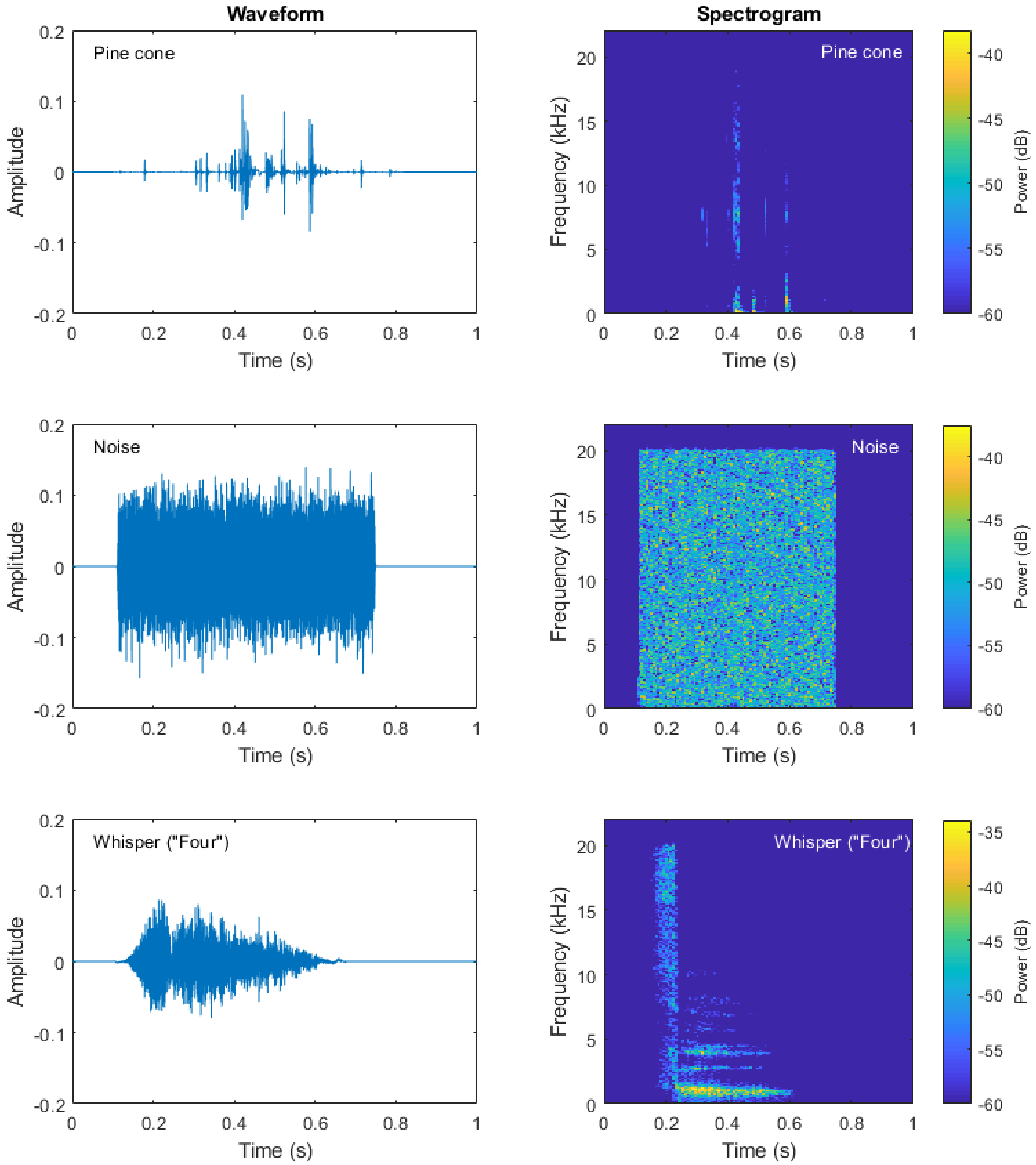

Figure 4.5 - Example waveform and spectrogram for each stimulus type. Each row shows a different stimulus type, pine cone, noise and whisper (the digit, "Four"); the left column shows the waveform and right column shows the spectrogram.

The final stage of stimulus processing was equalisation. A conventional way to equalise stimuli, such as speech, is by the RMS level of the signal. Due to the differences between stimuli in terms of their spectro-temporal characteristics, equalisation by RMS level could be problematic. The RMS level is likely to give an artificially low representation of the level for pine cones, due to the nature of the sound. As seen in Figure 4.5, the pine cone signal is essentially a series of transient clicks; as such, the peak level might be a better representation of the level of the pine cone, as it does not depend on the duration of the gaps in between peaks. However, the peak level could also be misleading, as one peak might be substantially higher than others, and subsequently misrepresent the signal.

The solution utilised to overcome this problem was the use of a novel 'target sensation level' model (TSL model), which combined a sound propagation model (ISO 9613-2) with a loudness model (Time-varying loudness model; (Moore et al., 2016)). The TSL model predicts the level at the subject location that is equivalent to the threshold level at the target location, such as in the situation shown in Figure 4.6.

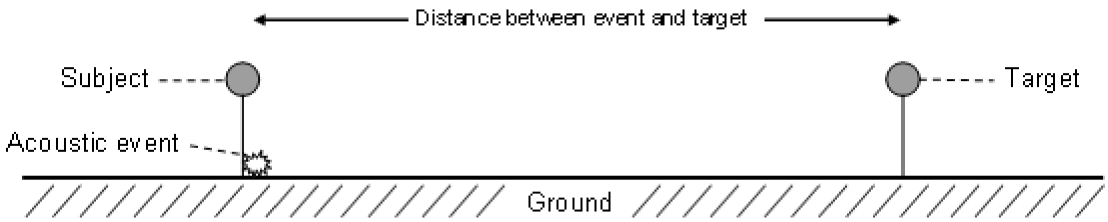

Figure 4.6 - Schematic representation of target and subject locations during aural detectability judgements.

This approach had the following advantages. Firstly, the time-varying loudness model utilised took into account the amplitude modulations and spectral differences present in each individual recording, so the problems related to the differing peak-to-RMS ratios and spectra could be overcome. Using a loudness model has been shown to be an effective approach for equalising the loudness of speech subjected to different signal processing that manipulated its dynamic fluctuations (Zorilă et al., 2016). Secondly, it allowed a pseudo-objective calculation of the aural detectability of sounds. This meant that the levels presented to the subject could be expressed as the sensation level at the target, meaning that an informed decision about the range of levels used in the experiment could be made. Thirdly, expressing the sound as a target sensation level helps the reader conceptualise the audibility of that sound in a particular condition. For example, if a broadband noise was presented to the subject at 60 dB SPL and the target was 50 m away, it is helpful for the reader/experimenter to know how that sound might be experienced by the target. The following sections describe the TSL model and how it was used to equalise stimuli.

4.4 Target sensation level model

4.4.1 Model description

A flow diagram showing each stage of the model is shown in Figure 4.7. The model uses the distance from source to subject and source to target, and the subject and target heights to first attenuate the sound as if it had travelled from the source to the target location. Then a loudness model calculates the peak short-term loudness experienced by the target. The model then adaptively varies the level of the sound at the subject in order to satisfy the threshold criterion loudness experienced by the target. The model returns the level at the subject location that is equivalent to the threshold level experienced by the target. This model was inspired by the model used in the aural non-detectability US military standard (R. Garinther et al., 1985).

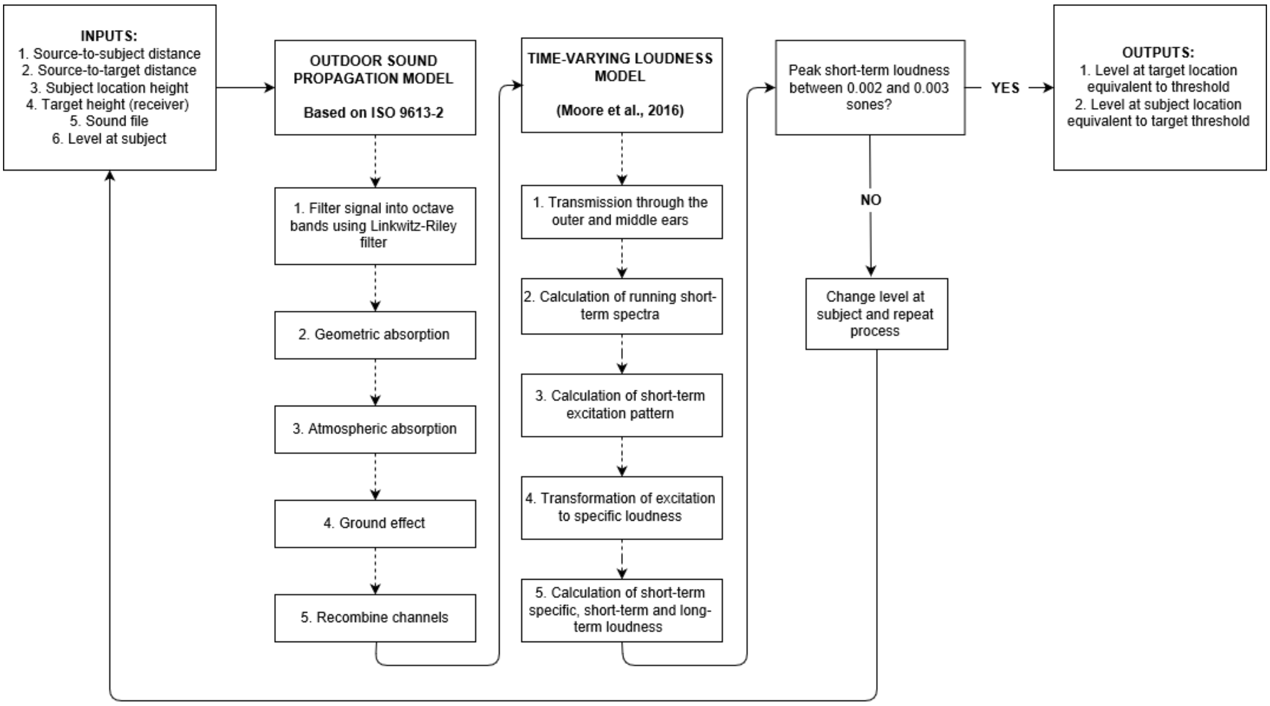

Figure 4.7 - Target sensation level model used to find level at subject equivalent to the threshold level at the target.

4.4.1.1 *Sound propagation model*

The sound propagation model used here is based on the ISO 9613-2 standard (Organization, 1996). The ISO standard calculates the level of a sound at an outdoor location, taking into account the sound source and various meteorological and non-meteorological factors. The amount of attenuation calculated by the model is linear in that it is independent of the input sound pressure level. The full model calculates the attenuation of a sound due to:

- Geometric divergence (inverse square law)
- Atmospheric absorption
- Ground effect
- Reflection from surfaces
- Screening by obstacles

Note, additional sources of attenuation occur in reality, including wind speed, wind direction and atmospheric temperature gradients. For simplicity, and due to the environmental conditions being simulated, these sources were ignored.

Due to the simple environment being simulated here, and to avoid unnecessary programming, a MATLAB implementation was created that modelled the relevant components of the standard. Because no surfaces (buildings, wall, etc.) or obstacles (foliage, trees, etc.) existed near or between the source and target, the effects of these sources of attenuation were disregarded, for ease of implementation. The MATLAB routine that was used to implement the sound propagation part of the model is shown in Appendix B.

The first step in the MATLAB routine filters the signal into octave bands, as the attenuation formulae are frequency dependent and applied in each octave band. The signal is filtered using a Linkwitz-Riley filter bank, which is comprised of a series of 2nd order Butterworth filters. Then, the attenuation due geometric divergence, atmospheric absorption and ground effects is applied to each frequency band.

The geometric divergence function accounts for the loss of energy due to the spherical spreading of the sound wave. The amount of attenuation (in dB) due to divergence is determined by the ratio of the source-to-reference (r_1) to source-to-receiver (r_2) distance, according to the following equation:

$$Attenuation_{\mathrm{div}} = 20\log_{10}\frac{r_1}{r_2}$$

This equation accounts for the inverse square law, where a doubling of distance from source to receiver results in 6 dB of attenuation.

The atmospheric absorption function takes into account the attenuation of sound via absorption in the atmosphere. The attenuation occurs due to shear viscosity, thermal conductivity, mass diffusion, thermal diffusion and the rotational relaxation of the molecules in air (Piercy et al., 1977). Atmospheric absorption is extremely frequency dependent, with high frequencies being attenuated more than low frequencies, so a frequency dependent function is applied to each band in the MATLAB function. The amount of attenuation is also dependent on the temperature and humidity of the atmosphere; for ecological validity, a relative humidity of 70% and temperature of 20 °C is used.

The ground effect function accounts for the effects of reflected sound off the ground between source and receiver and the direct path of sound between source and receiver. The source and receiver height, as well as the type of ground, are required for this calculation. A ground factor is used to calculate the effect of the ground type; porous ground was used in this function, as the environment used in experiment 3 was a grassy field.

Following the attenuation from the geometric divergence, atmospheric absorption and ground effects in each frequency band, the final step recombines the bands to give the signal at the receiver location.

4.4.1.2 *Time-varying loudness model*

The next stage of the TSL model is to calculate the peak short-term loudness experienced by the target. This is achieved by predicting the loudness using the time-varying loudness model developed by Moore et al. (2016). The MATLAB code used to implement the prediction of loudness was downloaded from the University of Cambridge website, https://www.psychol.cam.ac.uk/hearing/auditory-demonstrations-and-useful-software-1. The code was implemented according to the accompanying instructions. The loudness model is summarised in the following paragraphs. For a detailed review of the model and how it has been developed, the reader is directed to Moore (2014).

The loudness model accounts for binaural inhibition, so the following stages of the model are carried out for each ear independently. Note, binaural inhibition is not included in Figure 4.7 for simplicity. The model first filters the sound to account for the filtering effects during sound transmission through the outer and middle ear. The listening environment in the TSL model involves sounds presented in a free-field from a frontal incidence, so a free-field filter response is used here. Next, the model calculates the running short-term spectrum of the sound using 6 FFTs, each using different signal segment durations to give adequate spectral and temporal resolution across all frequencies. This is updated at 1-ms intervals.

The spectrum is then transformed into a short-term excitation pattern at 1 ms intervals, which represents the magnitude of the output of the auditory filters as a function of auditory filter centre frequency. The excitation pattern is then transformed into a specific loudness pattern, called instantaneous loudness; this is not a variable perceptible to the listener, rather, it is an intervening variable used in the calculation of loudness perception (Moore, 2014).

The next stage of the model calculates the short-term specific loudness pattern from a running average of the instantaneous loudness, as a function of centre frequency, in a process similar to that of an automatic gain control circuit. Then, a function calculates the smoothed and inhibited short-term specific loudness, which takes into account binaural inhibition. Finally, the model calculates the short-term loudness by summing the short-term loudness values for the two ears, and the long-term loudness by smoothing the short-term loudness for each ear, then summing the long-term loudness values for the two ears. The loudness is given in terms of phons or sones. The long-term loudness reflects the lasting impression of the sound, and is thought to involve higher-level processes, such as memory. The short-term loudness is more related to the loudness of a short segment, or part of a sound.

4.4.1.3 *Target threshold adaptive search*

The final stage of the TSL model is an adaptive process that determines whether the input sound (the sound level experienced by the subject) is at the threshold level of the target. For sounds at threshold levels, the peak short-term loudness of the sound should have a small but finite loudness, of approximately 0.003 sones (Glasberg and Moore, 2002). In the MATLAB implementation of the TSL model, the peak short-term loudness is considered to be at the threshold level for that sound if the value is between 0.002 and 0.003 sones. The following conditional statements were used in the adaptive procedure:

- if the value was greater than 0.5 sones, the input level was reduced by 10 dB
- if the value were greater than 0.05 sones, the input level was reduced by 5 dB
- if the value was greater than 0.01 sones, the input level was reduced by 2 dB
- if the value was between 0.003 and 0.01 sones, the input level was reduced by 0.5 dB
- if the value was less than 0.002 sones, the input level was increased by 0.5 dB
- if the value was between 0.002 and 0.003 sones, the adaptive procedure ended and the input level was equivalent to the target threshold.

The output of the model gives the level at the subject that is equivalent to the target threshold. This can be expressed either as the level experienced by the subject, *x* dB SPL (RMS or peak level), or as the sensation level experienced by the target, 0 dB SL. If a sound was presented from the same sound source, but 5 dB higher than the level determined to be equivalent to the target's threshold level, that sound could be expressed as *x* dB SPL + 5 dB at the subject, or 5 dB SL at the target.

4.4.2 Limitations

There are some limitations to the TSL model. Firstly, the loudness model assumes the listener (target) is responding with a moderate detectability index, *d'* (Moore, 2014). This may, or may not, be representative of the true detectability index employed by soldiers on watch duty or other situations where they might need to detect a threat. Secondly, the use of the loudness model for complex 'real-world' sounds such as whispers and pine cones may lead to inaccurate predictions due to a lack of validation for these sounds. The model has been extensively compared to synthetic stimuli, such as steady-state bands of noise and tones, to assess the accuracy of the model; these comparisons showed that the loudness model gives accurate predictions for these stimuli (Moore, 2014, Moore and Glasberg, 2007, Glasberg and Moore, 2006, Glasberg and Moore, 2002, Moore et al., 1997). Therefore, it is expected that the model will make accurate predictions for the noise stimuli used here. This could be problematic for the 'real-world' stimuli used as the stimuli used to tune the model were amplitude modulated tones and bands of noise (Moore et al., 1999). The applicability of the loudness model is further complicated due to the considerable variability across studies in the way that loudness changes with the duration of a sound (temporal integration of loudness) (Buus et al., 1997, Florentine et al., 1996), as the constants used in the loudness model equations cannot satisfy all published data. As such, it is possible that the loudness model may not give accurate predictions for all stimuli used here. Rennies et al. (2013) compared various loudness model predictions, including a previous version of loudness model used here (Glasberg and Moore, 2002), to loudness matches for speech-like and 'technical' sounds, and found that the loudness model used here had variable accuracy (0 – 15 dB error from mean behaviourally measured loudness matches). However, this experiment involved loudness matches at 70 dB SPL, so it is not possible to infer whether the variable accuracy would be present when predicting threshold levels for these sounds.

It was decided that the TSL model would be used to equalise stimuli due to the advantages of using the loudness model over RMS equalisation. For equalisation purposes, any errors associated with the loudness model should be consistent within a stimulus type. However, when using the TSL model for comparing participants' aural detectability judgements to the 'true' detectability, the potential inaccuracy of the model will be acknowledged.

4.4.3 Experimental stimulus equalisation

Prior to running the TSL model, a 40 dB SPL 1 kHz 1-s tone was generated in MATLAB and put through the loudness model. The output was checked for a peak short-term loudness level of 40 phons. This was done to ensure that the loudness model had been applied correctly.

The stimuli to be used in the pilot experiments (10 different pine cones, noises and whispers; 30 sounds in total) were equalised using the TSL model. Each sound was processed using the adaptive procedure described in the previous section to give the level at the subject that was equivalent to the threshold level at the target, when the target was 25, 50 and 100 m away from the sound source.

The levels at the subject, expressed as the RMS or peak level, that are equivalent to the target threshold level according to the TSL model, are summarised in Table 4-2. The mean and standard deviation are shown to summarise the levels for each condition.

Table 4-2. Summary of levels at subject equivalent to the predicted target threshold level, according to the TSL model. Means and standard deviations (SD) are given for each sound type, at each target distance, expressed as both RMS and peak level (dB SPL).

Target distance	**25 m**				**50 m**				**100 m**			
Level, dB SPL	RMS		Peak		RMS		Peak		RMS		Peak	
Statistic	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Whisper	25.5	2.0	41.8	1.4	32.0	2.0	48.3	1.4	38.8	2.0	55.0	1.2
Pine cone	15.8	2.8	44.2	1.5	22.3	2.8	50.7	1.5	29.3	2.8	57.7	1.5
Noise	33.2	0.4	45.5	0.7	39.8	0.2	52.1	0.5	46.9	0.3	59.2	0.5

4.5 Software and hardware development

This section describes the development of the software and hardware used in the pilot experiment, and subsequently for Experiment 2. The aim of this section is to describe what was done and how challenges were overcome, in order to give the reader a better understanding of the approach used here.

4.5.1 Software development

To develop the software required to run Experiment 2, the exact same virtual environment was used to that of Experiment 1. This environment was a simulation of an open grassy field located in Southampton Common. The same environment was used so that the distance estimation results obtained in Experiment 1, which showed that distance estimates were similar in reality and virtual reality, would be maintained in Experiment 2. Using the same environment also avoided the costly development of a new virtual environment.

Entirely novel code was written to run Experiment 2. The software was written to achieve the procedure shown in Figure 4.8. In summary, there are two stages: preparation and data collection. The preparation stage allows the researchers to input the experiment details (participant identifier, session number) and the experimental conditions to be tested. This stage was designed to permit easy manipulation for future experiments. Once the details had been inputted via text dialogue boxes and dropdown menus, the program generated the experimental trials list. The trials list was then randomised so that the trials would be presented in a random order.

Following completion of the preparation stage, data collection commenced. Each trial started with a large rectangle which obscured the view of the participant so they could not see the target whilst it changed position. The rectangle also displayed text that showed the trial number and the total number of trials, so the participant could monitor their progress in the experiment. The target was then positioned according to the trial list. The view-obscuring rectangle then faded out. Following a 1 – 2 s pause (randomised), which allowed the participant a moment to view the target, the sound was played. The participant then responded yes or no via a wireless controller (A or B button of an Xbox controller). Upon collecting the response, the software saved the data to a text file. The text file had a new line for each trial, and stored each trial parameter separated by a tab. A text file was used so the data could easily be imported into MATLAB for analysis.

As described in Section 4.3.2 and Figure 4.3, audio was presented via two loudspeakers. Loudspeaker 1 was located behind the participant and presented whisper and noise stimuli; loudspeaker 2 was located near the participant's feet and presented the pine cone stimuli. To ensure that the audio was produced faithfully and according to the procedure shown in Figure 4.8, the audio was played using the Unity audio toolbox. There are many audio features in Unity that are useful for gaming applications (e.g. compression, spatialisation); these were not useful for the current application, so all such features were turned off. Also, at the time of writing, there were limited available features for programmatically routing audio from specific sources within the game to specific channels of external soundcards. This limitation was overcome by programming the audio system as a stereo system, with one loudspeaker for each channel. Stereo audio files were created so that for noise and whisper stimuli, channel 1 contained the stimulus and channel 2 was 'blank' (just zeroes); for pine cones, channel 1 was 'blank' and channel 2 contained the stimulus. The audio toolbox worked by loading in the audio file relevant to that trial (stimulus type/level) and playing the file.

To accurately manipulate the level of the stimulus, the audio system in Unity was given a constant volume setting; the stimulus level manipulation was in the audio files. To calibrate the system, calibration audio files (20 seconds of 1 kHz tones at 40 and 80 dB SPL) for each loudspeaker were played and the gain setting on the loudspeakers was adjusted so that the correct levels were measured at the reference point by a sound level meter. The reference point for loudspeaker 1 (noise and whisper stimuli) was the position of the participant's head in the absence of the participant. The reference point for loudspeaker 2 (pine cone stimuli) was the microphone position as the pine cones were recorded (see section 4.3.3 for details).

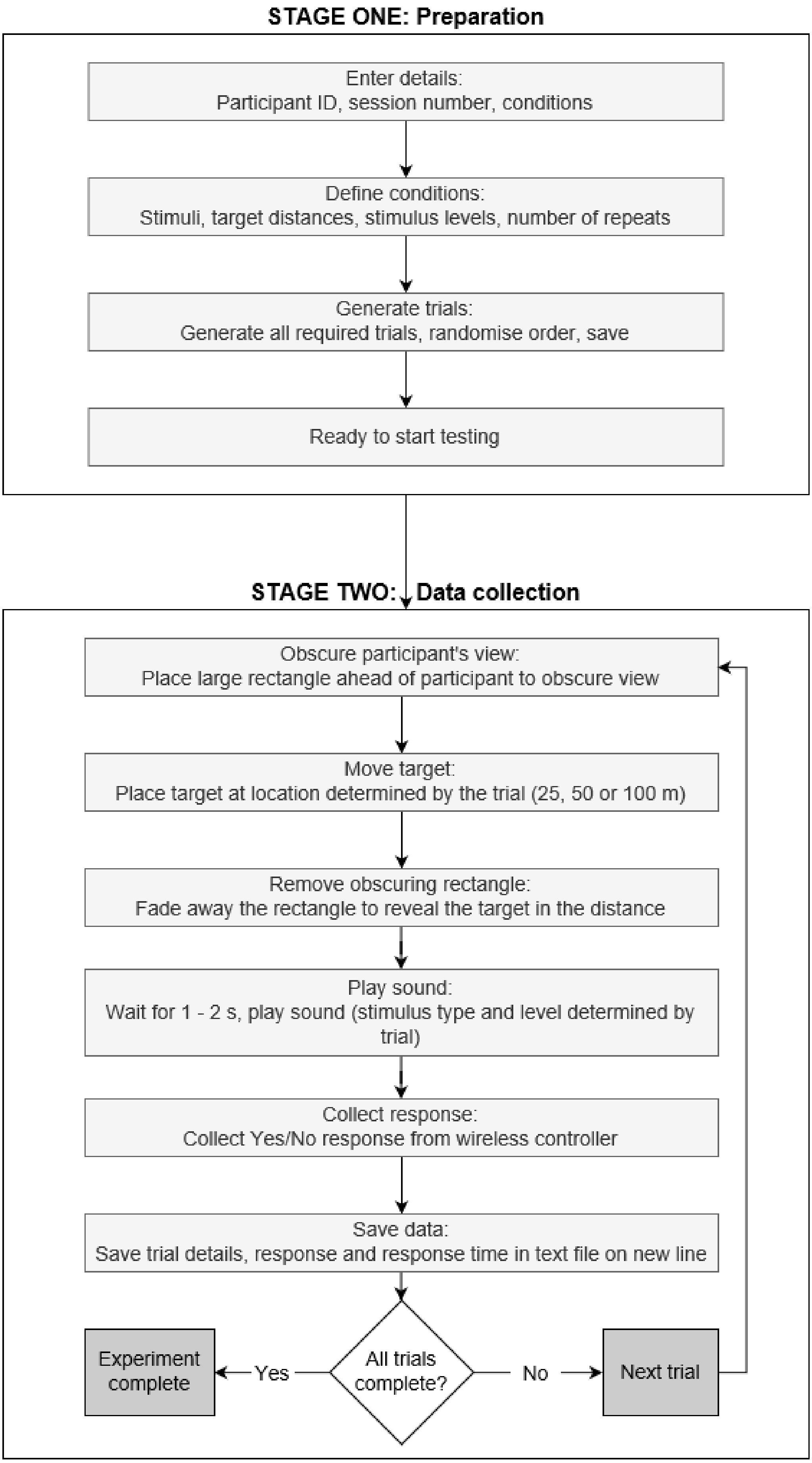

Figure 4.8 - Software procedure used in Experiment 2.

4.5.2 Screenshots

The following section shows some screenshots in order to give the reader an understanding of what the experiment was like for the participant. The software for Experiment 2 used the same ‘asymmetric game’ style as in Experiment 1, which allowed the researcher to view different information to the participant. Figure 4.9 shows the researcher’s view during the preparation stage, where the experimental conditions and participant’s details were inputted.

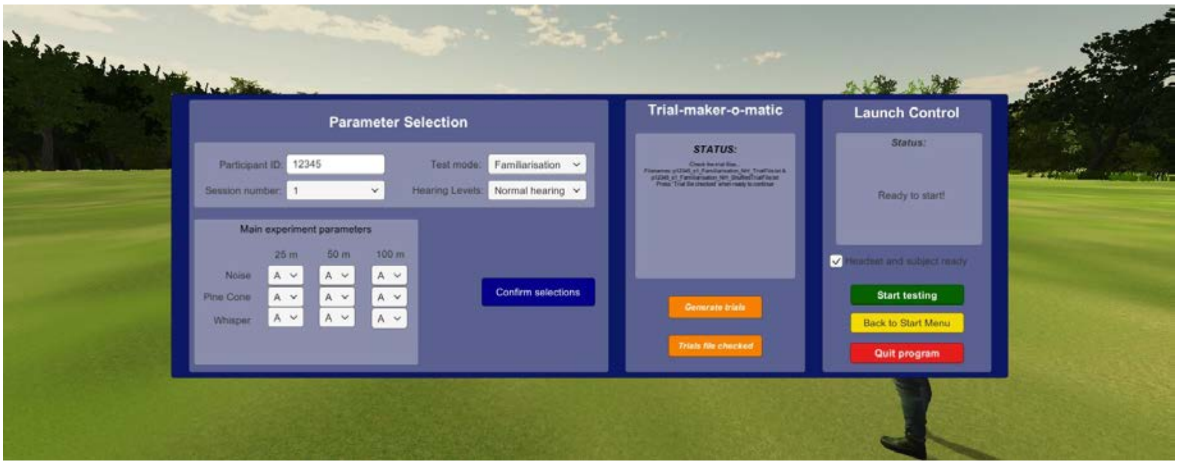

Figure 4.9 - Screenshot showing experimenter’s view during the preparation stage.

During the preparation stage, the participant view was obscured by a grey rectangle, shown in Figure 4.10. This grey screen was also present in between trials in order to obscure the view of the participant whilst the target was moved . The text on the grey screen also updated to show the trial number.

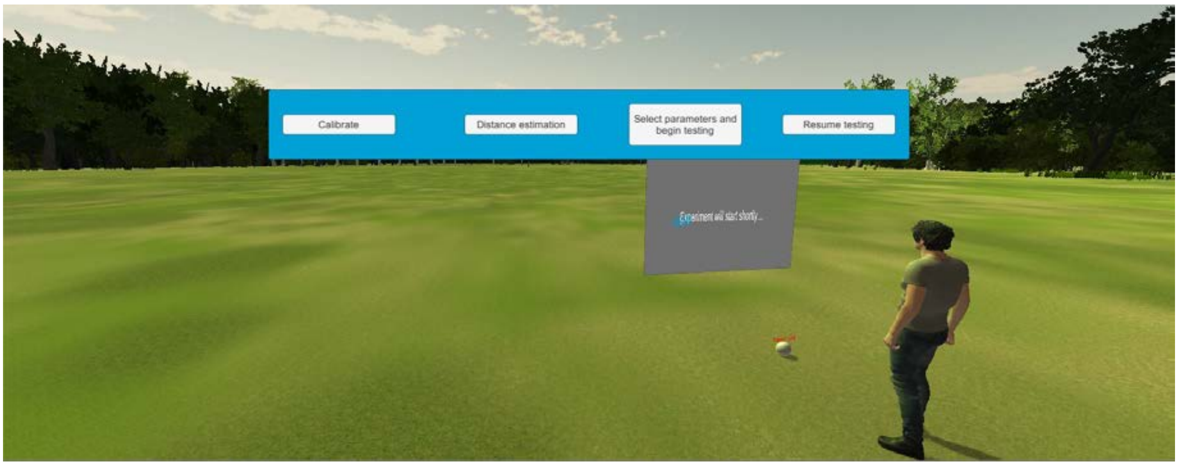

Figure 4.10 - Screenshot from experimenter’s view, showing grey screen that obscured the view of the target.

Figure 4.11 - Screenshot from experimenter's view of the scene during a trial.

During each trial, the grey screen is removed. This is shown in Figure 4.11. The human stood closest to the camera (person on the right of the screenshot) represents the person at the source location of the whisper; the loudspeaker was co-located with the virtual person's head. This screenshot depicts a trial with a target distance of 50 m. In the distance, the target can be seen. Note, the subject, represented by the sphere on the ground was not present when this screenshot was taken; the sphere would be seen at the approximate same height as the virtual person had someone been wearing the VR headset, as in Experiment 2.

4.5.3 Hardware development

The hardware required to achieve the methodology is shown in the schematic diagram in Figure 4.12. To simulate the anechoic conditions of an open field, the experiment takes place in an anechoic chamber. The small anechoic chamber at the University of Southampton was used. Whilst running the experiment, the laptop's fan must run quite noisily due to heat produced by the intensive processing. The laptop therefore cannot be in the same room as the subject, so it was routed out to a control room. A webcam was used so that the researcher could monitor the subject in the chamber.

As described previously, the subject was stood on a platform so that loudspeaker 2 could be placed close to the subject's foot, in order to reproduce the sound of stepping on a pine cone. The platform was made-to-measure in the University workshop; the frame was made of steel and the platform was Glass Reinforced Plastic (GRP) grating. The GRP grating was used because of its lightweight and non-slip properties. A photograph of a participant completing the experiment during the pilot experiment in shown in Figure 4.13.

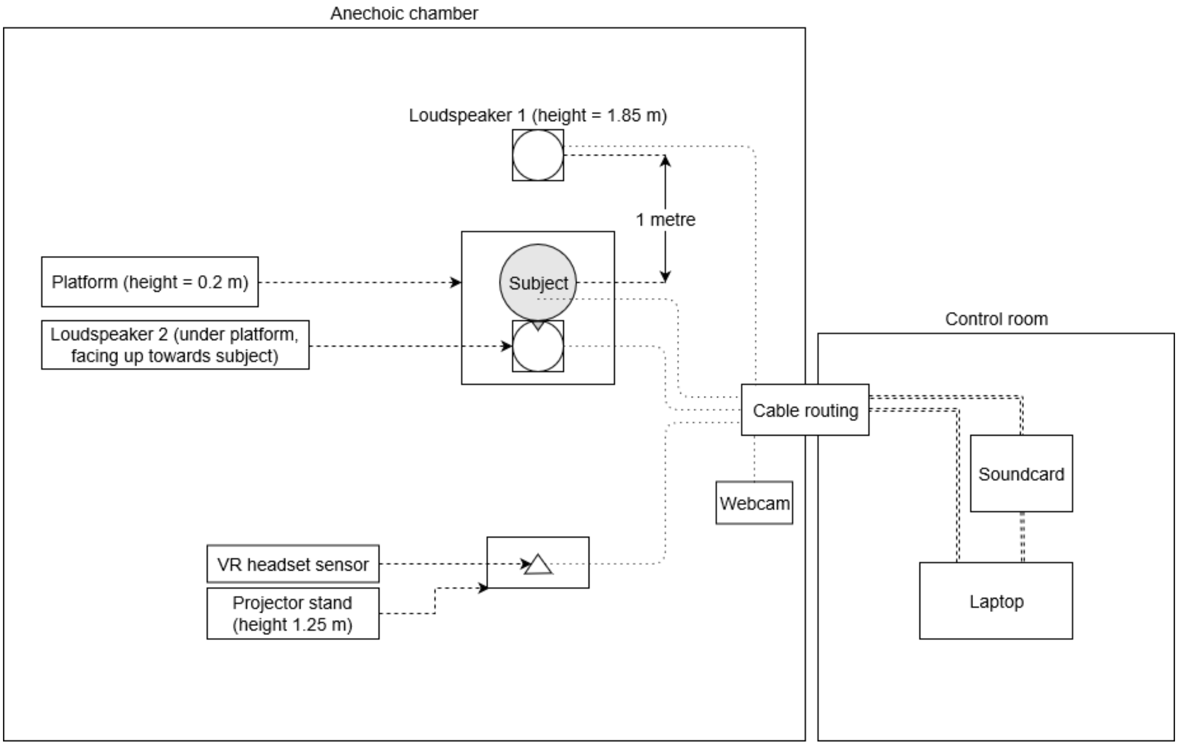

Figure 4.12 - Schematic layout of apparatus in experiment 2 and pilot experiments.

Figure 4.13 - Participant completing pilot experiment.

4.6 Pilot experiment

4.6.1 Introduction and summary of methods

The developed method was tested in a pilot experiment. The aim of the pilot experiment was to assess:

- Time constraints
- Conditions to use in main experiment (levels, target distances, number of repeats)
- Participant experience of the experiment.

The method used in the pilot experiment is summarised in Table 4-3. For a detailed discussion of each methodological factor, see the relevant previous section. The stimulus levels were chosen to give the participant a large range (45 dB) of levels in order to maximise the chances of measuring the complete psychometric function. It was predicted that a stimulus level of – 15 dB SL (predicted target sensation level), a sound equivalent to 15 dB below the target's threshold, would result in the subject reporting that that sound was not audible to the target. Likewise, it was predicted that 30 dB SL (predicted target sensation level), a sound equivalent to 30 dB above the target threshold, would result in the subject reporting that that sound was audible.

Table 4-3 - Summary of methodological factors used in pilot testing.

Methodological factor	Description	Relevant thesis section for further detail
Experimental task	Aural detectability judgement: subject views target, then listens to sound, then makes yes/no judgement whether the target would be able to aurally detect that sound	2.3.5
Psychoacoustic method	Method of constant stimuli, with 6 observations (stimulus levels) and 8 trials per observation.	2.4.4, 4.2
Independent variables	Stimulus type, target distance	2.4.5, see Stimulus types and Target distances.
Dependent variable	Proportion of 'Yes' responses at each level	4.2
Stimuli	Noise, Twig, Pine cone, Whisper	4.3
Target distances	25, 50, 75 m	4.5.1, 0
Levels (predicted target sensation level, dB)	-15, -5, 0, 5, 15, 30	4.4
Total number of trials	576 (3 distances x 4 stimuli x 6 levels x 8 repeats per level)	n/a
Apparatus	As described in section 4.5.	4.5
Task instructions	Given verbally	n/a
Number of participants	n = 4 (n = 3 in noise condition)	n/a

4.6.2 Results

The results of the pilot experiment are shown in Figure 4.14. Each plot shows the proportion of 'Yes' responses at each stimulus level for each individual (n = 4). Due to a technical fault, one subject did not complete testing for the noise stimulus type.

The reader is reminded that the stimulus level expressed in terms of the predicted target sensation level shows how that particular stimulus would have been perceived by the target in that particular condition; 0 dB SL means the sound that was presented to the subject was at a level equivalent to the target's threshold. A stimulus level of 30 dB SL means the sound would have been received by the target at 30 dB above their threshold.

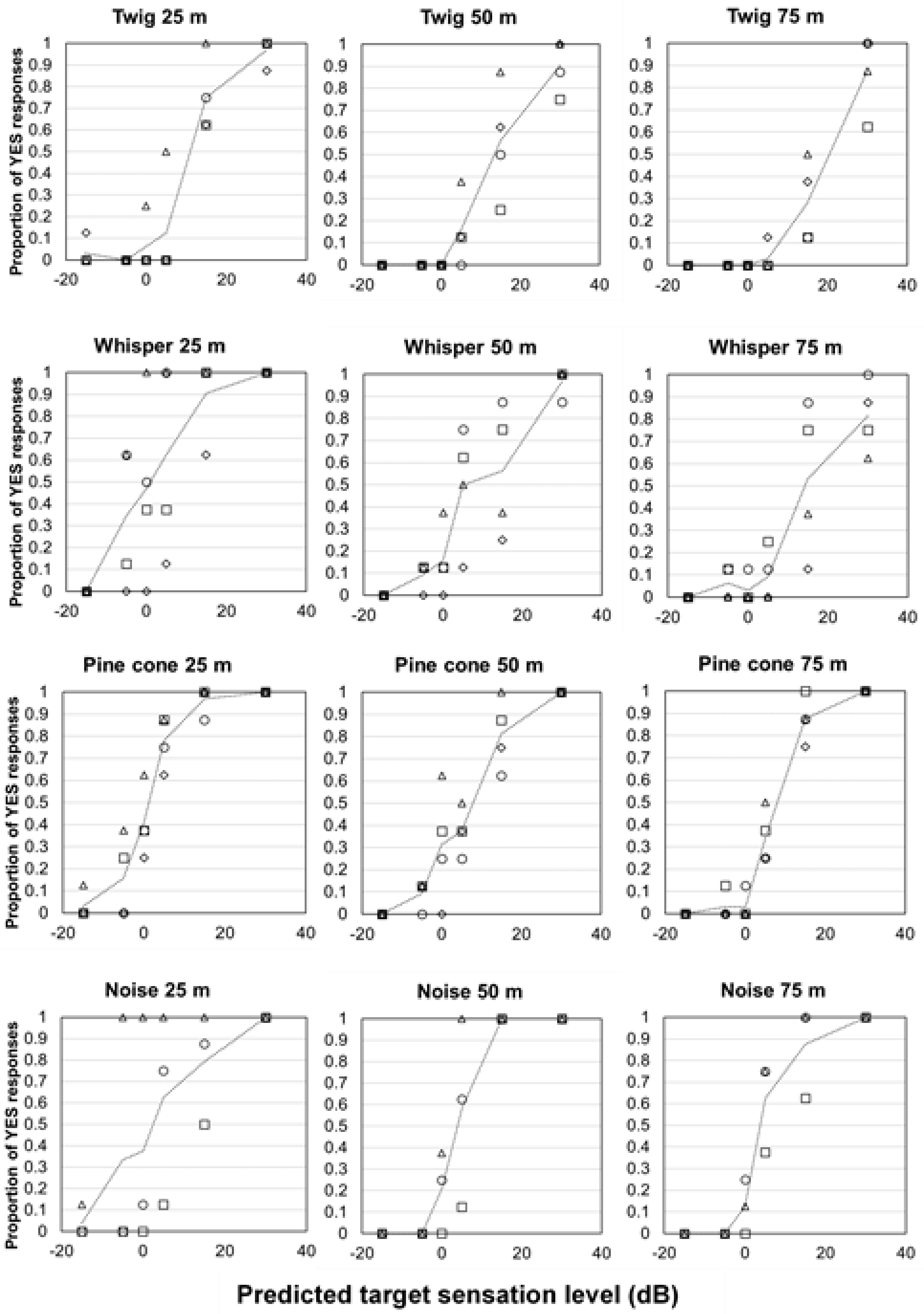

Figure 4.14 - Results of pilot experiment for each stimulus type and target distance. The abscissa shows the stimulus level in terms of the predicted sensation level experienced by the target, where 0 dB is equivalent to the target's threshold for that sound. The ordinate shows the proportion of 'Yes' responses for that stimulus level. Each symbol shows a different participant (n = 4). The line connects the mean judgement.

4.6.3 Discussion and recommendations

The participants who took part in the pilot study tolerated testing well. Towards the end of the testing session they felt somewhat fatigued, but at no point did they feel uncomfortable and want to stop. The participants also reported preferring frequent short breaks, every 150 - 200 trials. Each session took approximately 1.25 – 1.75 hours; each trial lasted 8 – 9 seconds. This length of testing was comfortable for the participant, so the main experiment should aim be of similar length.

The subjects reported no difficulty understanding the experimental task. The instructions were delivered verbally which may result in some inconsistency. Therefore, the instructions should be given in a written format during the main experiment.

The PF transitions from p(yes) = 0 to p(yes) = 1, with the participant judging the sound to be inaudible to the target at relatively low levels and audible at relatively high levels. This shows that the method used here has, to some extent, construct validity, as the stimulus level must modulate the p(yes) responses if the participant is truly judging the aural detectability of the sound.

There was quite large variability between subjects. Using a wide range of stimulus levels (-15 to 30 dB SL) allowed this to be measured, but at the cost of wasted trials. There were several conditions where a subject had p(yes) measurements at 1 or 0 at multiple levels; this is inefficient testing as a large number of trials are spent measuring the asymptotes of the function, rather than the threshold portion of the function (see Noise 25m and 50m triangles, Whisper conditions diamonds, Twig conditions squares). It would also be preferable to have measurements with reduced spacing, in order to measure the PF with more resolution.

A solution to this is to not use the same observation levels (stimulus levels) for each participant. This would involve briefly obtaining an initial estimate of the PF at the start of the testing session, then selecting a set of stimulus levels for the remaining experimental trials based on the initial estimate of the PF. The set of stimulus levels for the main experiment could then be spaced more closely (than the spacing used in the pilot experiment), allowing measurement of the PF with more precision. The aim of the initial estimate of the PF would be to estimate the location of the PF, rather than accurate and precise mapping of the full function.

The method by which the initial estimate of the PF is measured could be similar to that used in the pilot experiment (method of constants), or an alternative method such as the method of adjustment. Using the method of constants for the initial estimate of the PF means the subject experiences the same task throughout the whole experiment, so they do not have to learn two sets of task instructions. They also gain task familiarity by practising the task prior to the main experimental trials. It is also easy to implement in terms of the software, as a new method does not need to be programmed. There are however disadvantages. Using the method of constants is not the most rapid psychoacoustic method for estimating the location of a PF (Gescheider, 2013). The number of trials per stimulus level would have to be reduced substantially in order to speed up the measurement, which risks greater measurement error due to reduced number of repeats. Based on the advantages of task familiarity and ease of implementation, and on the balance of risk associated with a limited number of repeats per stimulus level, it was decided that the method of constants would be used to obtain an initial estimate of the location of the PF. This would take place at the start of the testing session, and will hereon be called the familiarisation period.

The inclusion of a familiarisation period increased the testing time per session. Given the time constraints expressed by subjects in order to avoid discomfort or fatigue (maximum trials per session = 550-600), the number of conditions for the main experiment needed to be reduced. It was decided that the twig stimulus would be removed. By keeping the noise, whisper and pine cone, the effect of stimulus familiarity could still be assessed, and both a locomotion and communication related sound was maintained for ecological validity. By keeping the number of target distances at 3, the total number of conditions was reduced from 12 to 9.

It was decided that the familiarisation period would include a wide range of stimulus levels per condition, in order to maximise the chance of measuring the upper and lower asymptote of the participants PF. Based on the results of the pilot experiment, it was decided that the following 6 levels would be used: -18, -8, 2, 12, 22, 32 dB SL. There would be 2 repeats at each stimulus level, resulting in 12 trials per conditions. For the familiarisation period, the total number of trials would be 108 (9 conditions x 12 trials per condition). For the remaining experimental trials, it was decided that 6 stimulus levels would be used, but the number of repeats at each level would be increased from 8 to 10, in order to increase the precision of the p(yes) measurement. This results in 540 trials for the main experiment (9 conditions x 60 trials per condition). In total, there were 648 trials. To avoid fatigue, the participant was given a break after the familiarisation period, and after trial 180 and 360 in the main block of trials.

Another observation from the results of the pilot experiment is that there appears to be a trend for the PF to increase in stimulus level (it moves to the right on the x-axis) with increasing distance. Since experiment 2 showed that distance estimation in VR was similar to that for reality at 100 m, it was decided that the 75 m target distance would be changed to 100 m, in order to maximise the effect of target distance. This is also conceptually easier to interpret, as a doubling of distance results in an approximate change of 6 dB in the actual aural detectability of the sound.

4.6.4 Summary

In summary, the pilot experiment demonstrated that the method was able to measure the psychometric function and the participants tolerated the testing well. It was decided that a familiarisation period, which helps determine the set of stimulus levels for the main block of trials would be used. The 75 m target distance condition was changed to 100 m.

Chapter 5 Aural detectability judgements by normal-hearing civilians

5.1 Experiment 2: Introduction

This chapter reports experiment 2, which measured aural detectability judgements by normal-hearing civilians using the methods developed and discussed in Chapter 4. A summarised version of the methods is given to show the exact conditions tested; the reader is referred to Chapter 4 for a thorough documentation and justification of the various aspects in the method.

5.2 Aims and predictions

The objective of this experiment was to measure aural detectability judgements by normal-hearing civilian subjects. The aims of the experiment were to test whether judgements are:

- Repeatable
- Sensitive to stimulus level
- Sensitive to target distance (distance between subject and target)
- Affected by the type of stimulus
- Accurate (judgement matches the true detectability).

The following sections summarise the predictions for each of these. For a full discussion of each of these factors, see section 2.4.5.

5.2.1 Test-retest reliability

It was hypothesised that if subjects are making judgements based on consistent decision criteria, aural detectability judgements should be similar when measured on separate and independent occasions. To test the repeatability of judgements, two sessions were included in the experiment. The second session was a repeat of the first session. It was predicted that the judgements in session 2 would be similar to those for session 1 as there is no reason to suspect that a systematic change would occur.

5.2.2 Sensitivity to stimulus level

The detectability of a sound depends on its level relative to a listener's auditory threshold. Therefore, aural detectability judgements should depend on the level of the stimulus as experienced by the subject. If the subject's judgements are sensitive to the level of the stimulus, it shows that aural detectability judgements do depend on stimulus level, and, to some extent, that the method used here has reasonable construct validity (the ability of a test to measure what it purports to measure (Cronbach and Meehl, 1955)). It was shown in the pilot study (Section 0) that higher stimulus levels resulted in a higher proportion of 'yes' responses, which suggests that aural detectability judgements are sensitive to stimulus level. It was predicted that this would occur for a larger sample in each condition tested. Stimulus level was explored here by virtue of the psychoacoustic method; the method of constants was used which allows an analysis of how judgements change across different stimulus levels, within a single condition.

5.2.3 Sensitivity to target distance

Intensity is known to be a cue for egocentric auditory distance perception (Kolarik et al., 2016). Therefore, it was predicted that subjects would incorporate sound propagation into their aural detectability judgements. Whether they judge sound attenuation rates correctly is more difficult to predict and measure. There are data to show that people can make accurate vocalisation level adjustments (~6 dB per doubling of distance) to compensate for increased distance to target (Zahorik and Kelly, 2007); however, only distances up to 8 m were tested and they used a different experimental task, and therefore no specific predictions can be made for the current experiment.

The effect of target distance was explored using three target distances: 25, 50 and 100 m. These were selected because they are representative of the proximity that a soldier might have to function in during stealth operations. Also, distance estimation in virtual reality was verified to be similar to that in reality on average over these distances, which helps to maintain the construct validity of the method used here.

5.2.4 Stimulus type

This project was primarily motivated by a practical interest in the accuracy with which soldiers can judge whether the noise they make during an operation can be heard by an enemy some distance away. Consequently, this experiment used two sounds that were considered to be ecologically relevant to situations where soldiers might need to make such judgements (e.g. whilst moving or communicating). These were:

- Pine cone (the sound of stepping on a pine cone)
- Whisper (a whispered digit)

One reason that judgements might vary between these two sounds is that they have different spectro-temporal features. For example, the whisper sound contains more high-frequency energy than the pine cone; the peak-to-RMS ratio is greater for the pine cone than for the whisper (see Section 4.3.4). By using these two ecologically relevant stimuli only, it is not possible to assess how specific spectro-temporal features affect aural detectability judgements. However, the use of ecologically valid sounds was more of a priority at this early stage of ASA research. Future research should investigate the specific spectro-temporal properties that might affect aural detectability judgements.

Another reason the type of stimulus might affect aural detectability judgements is that prior experience of the stimulus might bias judgements (stimulus familiarity). For example, the experience of listeners that whispers are usually quiet, or perceived with the whisperer nearby, might influence their judgements of whether a target some distance away could hear them, even at sufficiently high intensities for the whispers to be heard. Studies have shown that vocal effort affects auditory distance perception estimates for equal presentation levels (Kolarik et al., 2016, Brungart and Scott, 2001, Philbeck and Mershon, 2002), which suggests that factors in addition to the spectro-temporal properties and intensity of sounds could affect aural detectability judgements.

Given the potential for a role of stimulus familiarity, the effects of stimulus levels and distance to provide information on the construct validity of the measurement technique could be compromised with the ecologically relevant sounds (pine cone and whisper). A third stimulus that seemed less likely to be influenced by familiarity was therefore included; it is also a stimulus that has been commonly used in studies of auditory perception:

- Noise (‘synthetic’ Gaussian noise)

The noise sound was included to explore whether prior experience of a sound aided the accuracy of aural detectability judgements.

As discussed in Section 2.4.5, it was predicted that due to the prior experience of pine cones and whispered speech would aid the accuracy of the judgements. Therefore, it was predicted that judgements for whispers and pine cones would be more accurate than noise, that is, closer to the modelled correct aural detectability.

5.2.5 Accuracy of judgements

Accuracy is defined here as the difference between the participant’s judgement and the true detectability of the sound. The accuracy is measured here by comparing the participant’s judgements to predictions of the target sensation level model (Section 4.4).

Predictions about the accuracy of judgements were difficult to make for each stimulus type and target distance due to the exploratory nature of this study. However, it was expected that participants would make broadly accurate judgements, based on the hypothesis that prior experience of sounds would inform judgements.

5.3 Methods

5.3.1 Participants

Thirty young normal-hearing civilian participants took part in the experiment. Two participants dropped out after the first session voluntarily, leaving a final sample size of 28. Normal-hearing was tested for using pure tone audiometry; thresholds less than 20 dB HL between 250 and 8000 Hz were considered normal hearing. A screening questionnaire that checked for age, vision problems, ear disease, tinnitus and recent noise exposure was conducted to ensure that each participant was aged between 18 – 40 years, had healthy ears and had normal, or corrected-to-normal vision. Participants were all undergraduate or postgraduate students at the University of Southampton.

5.3.2 Apparatus

The experiment took place in an anechoic chamber. The schematic diagram in Figure 5.1 shows the layout of the testing environment and apparatus used. The subject stood throughout the experiment on a custom made platform, with a Glass Reinforced Plastic (GRP) grating. Underneath the platform was a loudspeaker that produced the pine cone stimuli (Loudspeaker 2). The loudspeaker cone was 10 cm ahead of the subject centred between their feet. It was angled upwards, facing directly at the subject's head. Small wooden slats were placed on top of the grating to stop the subject from moving their feet and covering the loudspeaker.

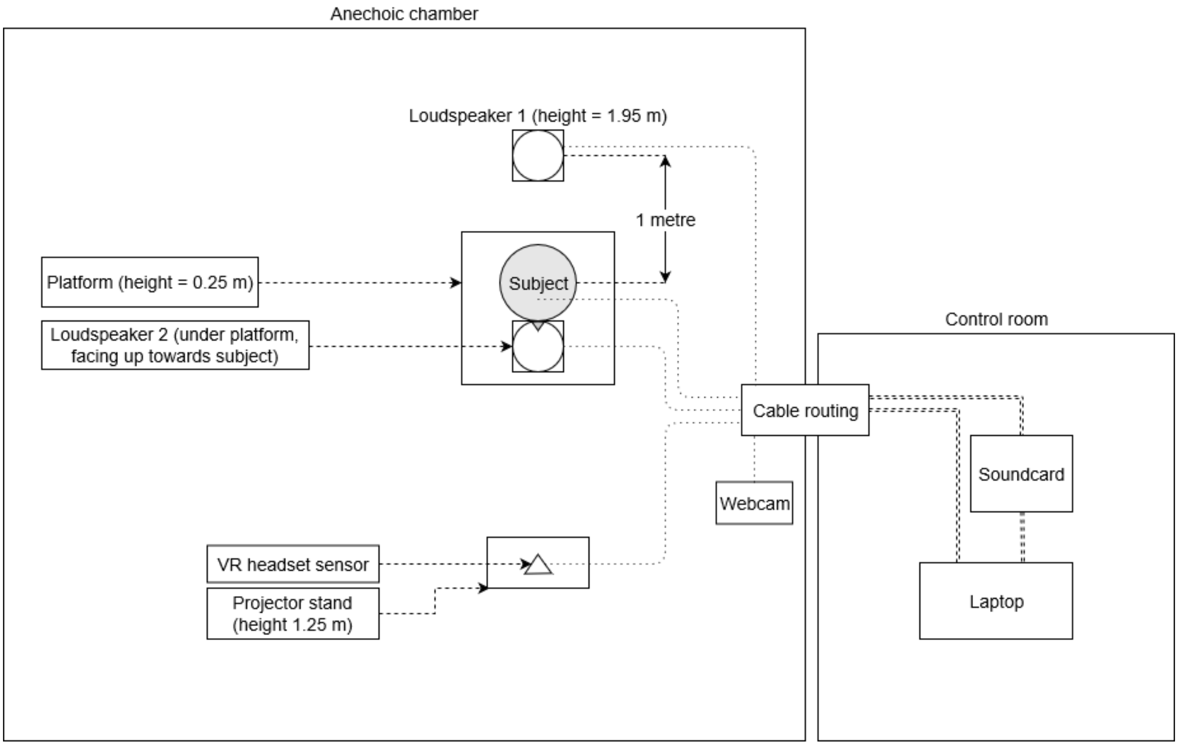

Figure 5.1 - Schematic diagram of Experiment 2 set up.

The other loudspeaker was placed at a height of 1.95 m for all subjects (Loudspeaker 1). This location approximately matched the position of the virtual talker's mouth, whose height was 1.75 m. This loudspeaker produced the noise and whisper stimuli.

The virtual reality headset was worn by the subject and the sensor was placed in front of the subject on a projector platform. The virtual environment that the subject viewed was described in Section4.5.2. All cables were routed out of the chamber into the control room, where the researcher observed the subject via the webcam.

5.3.3 Conditions

There were 3 target distances (25, 50 and 100 m) and 3 stimulus types (Pine cone, Noise and Whisper), resulting in 9 conditions. In each condition, the subject was presented with 6 different stimulus levels, as part of the method of constant stimuli. All conditions were repeated during the second testing session.

5.3.4 Stimuli and calibration

For a full description of the stimuli used in this experiment, see Section 4.2 and 4.4. In short, for each condition and at each of the 6 stimulus levels, the subject was presented with 10 different recordings (and 10 different noise waveforms). The proportion of those 10 presentations to which the subject responded 'yes' was the outcome measure (p(yes)). Note, the same 10 recordings for each stimulus were used in each condition at each distance.

The six stimulus levels for each condition in the main testing were determined during a familiarisation period for each subject individually (see pilot study in Section 0 for justification). Familiarisation testing consisted of the same set of six levels for each condition for all subjects. The set of stimulus levels used in the main experiment were then determined based on how the subject responded in the familiarisation period. One of four sets of stimulus levels was selected for each condition (set A, B, C or D). The set was selected such that p(yes) = 0.5 appeared to be roughly in the middle of the set of stimuli. This selection took place whilst the subject rested in between the familiarisation period and the main testing.

Due to the equalisation process, discussed in detail in Section 4.4.3, stimulus level could be represented in terms of the RMS level experienced by the participant, or the predicted target sensation level (dB) (according to the model discussed in Section 4.4). When represented as the predicted target sensation level, the levels for each set are as shown in Table 5-1. These levels expressed as the RMS level (in dB SPL) as experienced by the subject for all stimuli are shown in Appendix D, for reference.

Table 5-1 - Stimulus levels, expressed as predicted target sensation level (dB) for each set of stimulus levels in Experiment 2.

Set name	**Levels (predicted target sensation level, dB)**					
	1	***2***	***3***	***4***	***5***	***6***
Familiarisation	-18	-8	2	12	22	32
A	-18	-9	-5	-1	3	12
B	-8	1	5	9	13	22
C	2	11	15	19	23	32
D	12	21	25	29	33	42

The loudspeakers were calibrated by presenting a 1 kHz tone at 80 dB SPL and 40 dB SPL and measuring the level at the reference point with a calibrated sound level meter. The loudspeaker gain control was adjusted so that the level at the reference point was as intended. 40 and 80 dB SPL tones were used to check that the loudspeaker faithfully reproduced low and high levels.

For Loudspeaker 1 (Noise and Whisper stimuli; behind the subject's head), the reference point was the location of the subject's head with the subject absent. For Loudspeaker 2 (Pine cone stimuli; underneath the platform), the reference point was 1.75 m from the loudspeaker (head height), slightly ahead of the subject's head position, with the subject stood in place. This meant the reference point was effectively approximately 40 cm in front of the subject's head. This was done to preserve the transfer function between the source and subject, as the subject was present whilst recording the pine cone sounds. The reference point used in this experiment was the same as that used for recording. For full details and explanation, see Section 4.3.3.

For the analyses of the effects of session, target distance, stimulus level and stimulus type, the level of the stimulus was expressed as the RMS level experienced by the subject. This was because the hypotheses related to these experimental variables were concerned with how the stimulus is perceived by the subject, not the target. For the analysis of the accuracy of judgements, the level of the stimulus was expressed in terms of the predicted target sensation level.

5.3.5 Procedure

The experiment was conducted according to the following procedure. First the participant attended a screening session to check that they were eligible for the experiment. The subject also completed the consent form. If eligible and consent was given, the subject was then invited to attend the two experimental sessions.

The two experimental sessions were identical. First the subject read the following written instructions:

Task instructions

Each trial starts with a grey screen blocking your view. It will then fade out, and you will see a person standing in the distance. Then, you will hear a sound coming from either your feet or a person stood behind you. The sounds behind you will be either a whisper or a noise. The sounds from your feet will be as if you've just stepped on a pine cone. You must then decide, yes or no, would the person in the distance be able to detect that sound?

Importantly, it is not whether or not you can hear the sound, it is whether you think the person in the distance would be able to hear the sound.

You will give your judgement by pressing either the 'A' or 'B' button on the controller.

If you decide 'Yes, the person would be able to hear that sound', press A.

If you decide 'No, the person would not be able to hear that sound', press B.

Once you have given your judgement, the trial will finish and we will move onto the next trial.

The person will change location throughout the experiment and the sound will change.

You should assume that the person has normal hearing and is actively listening for the sound. They cannot see the sound being created, so you should not factor sight into your decision.

The visual and sound environment is exactly as you are experiencing. It is extremely quiet, there is no background noise and it is a very calm day with no wind.

Safety

Try to move as little as possible during the experiment. Be aware of the small platform you are standing on. Please do not stand on the wooden slats that are designed to stop you from putting your foot over the loudspeaker.

If for any reason you want to stop or break, please inform the experimenter by waving your hand. If you feel dizzy or uncomfortable, please inform the experimenter.

For your comfort, we will take a few breaks throughout the experiment.

Practice

We will start with some practice trials. These trials will not count towards your results.

Reminder: you will see a person in the distance, then you will hear a sound, you need to decide yes or no, would they be able to detect that sound?

Once the participant had read the instructions, they were invited to ask questions. Their understanding of the instructions was then checked.

Following task instructions, the participant was taken into the anechoic chamber and had the VR headset fitted so that it was comfortable and secure. The headset was then calibrated using the Oculus Rift headset calibration software, which ensured that the position of the headset and the inter-pupillary distance were correct. The virtual environment was then displayed. The researcher encouraged the participant to look around the environment to let them become immersed in the display. To help calibrate their distance perception, the subject was shown the target at 10 m away from them. The researcher told the subject that the target was 10 m away. This was identical to the calibration procedure used in Experiment 1. The subject was told that at the end of the experiment they would be asked to estimate the distance between themselves and the target at each target location, to the nearest metre.

Then, a familiarisation period commenced. Every subject experienced the same 108 familiarisation trials, though a different random order was used for each subject. These trials were made up of each target distance and stimulus (9 conditions); there were 6 stimulus levels for each condition; each condition and stimulus level was repeated, resulting in 108 trials.

Each trial in the familiarisation period and main experiment followed the procedure shown in Figure 5.2. In the familiarisation period, subjects were not given a break. It took approximately 10 – 14 minutes.

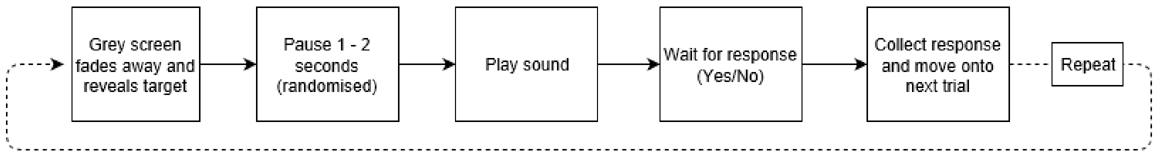

Figure 5.2 - Aural detectability judgement experimental task procedure.

Following the familiarisation period, the subject removed the headset and had a break whilst the researcher configured the main experimental trials. The set of stimulus levels was selected in order to maximise the chance of capturing the transition between p(yes) = 0 and 1.For subjects for whom the familiarisation period did not result in a p(yes) = 1 measurement, the highest set of stimulus levels (D) was selected.

The main experimental period then commenced. For each condition (9) and each stimulus level (6 per condition) there were 10 sounds. This resulted in 540 trials. The 540 trials were presented in a random order. Subjects were given breaks after trials 180 and 360. At the end of the trials, the participant was asked to estimate the distance between themselves and the target when the target was positioned at each of the target locations in the main experiment (25, 50 and 100 m). Immediately prior to making these estimates, the subject was shown the target at 10 m and told that it was at 10 m. A grey screen them blocked the view of the subject whilst the target was moved to the first target location. The subject then verbally reported their distance estimate. This was repeated for the remaining target distances. Target distances were presented in a randomised order.

The second session was identical to the first session. Subjects were asked to read the written instructions again and the full familiarisation period was repeated. The same selection process for choosing the sets of stimulus level was used in the second session, so that the researcher did not introduce bias by influencing the stimulus levels that the participant was presented with.

Each session took between 1.5 and 2 hours, depending on how quickly the participant responded on each trial. Upon completion of both sessions, the participant was paid £60.

5.3.6 Analysis

5.3.6.1 *Planned analysis*

The analysis is broken down into two sections. The first explores the effect of stimulus level on judgements, reports the fitting of psychometric functions, checks assumptions and removes outliers. The second analyses the locations and slopes of psychometric functions in terms of the test-retest reliability, the effects of target distance and stimulus type, and the accuracy of judgements.

Psychometric functions (PFs) were fitted using MATLAB routines that estimated the parameters of the PF and how well the fit describes the data. Our laboratory is most familiar with the Palamedes Toolbox implementation (Prins and Kingdom, 2018), so this approach was used. See the following sections for a fuller explanation of how PF parameters were estimated.

The planned statistical analyses were repeated measures ANOVAs on the locations and slopes of the PFs, with session, stimulus type and target distance as factors. This produced three main effects and four interactions per ANOVA. The accuracy of judgements was explored by expressing the PF locations in terms of the predicted target sensation level.

As described in the Stimuli section of the methods for this experiment, the stimulus level can be expressed as either the RMS level of the sound experienced by the subject, or the predicted target sensation level according to the target sensation level model.

5.3.6.2 *Fitting psychometric functions*

The PFs were fit using a maximum likelihood method to estimate the parameters of the PF with the following equation:

$$\psi(x;\alpha,\beta,\gamma,\lambda) = \gamma + (1-\gamma-\lambda)F(x;\alpha,\beta)$$

The function F(x;α,β) was a logistic function, as it was expected that a sigmoidal function with asymptotes at p(yes) = 0 and p(yes) = 1 would likely suffice for characterising the data here. The parameters that best describe the psychometric function were estimated using MATLAB routines, and described the data using the following equation:

$$F(x;\alpha,\beta) = \frac{1}{1+\exp(-\beta(x-\alpha))}$$

Parameter α corresponds to the location of the PF, which represents the stimulus magnitude level of x equivalent to the midpoint of the PF: $F(x=\alpha;\alpha,\beta)$ = 0.5. As α represents a stimulus magnitude, it has the units of the stimulus; in this study, the unit is the sound pressure level of the stimulus, in dB SPL. Parameter β determines the slope of the PF, though its value does not directly respond to the slope of the function as defined in calculus (Wichmann and Hill, 2001). Rather, it is a dimensionless parameter associated with the logistic function[6]. A higher number represents a steeper slope. The slope of the PF will not be converted here as this is the first study of its kind and therefore converting the estimate to a parameter comparable between studies is not necessary (a conversion would be required when, for example, reviewing the slope of PFs for speech in noise perception across various studies, e.g. MacPherson and Akeroyd (2014)). Parameter γ is the guess rate, which corresponds to the probability of a 'yes' response when the stimulus is below their criterion. Parameter λ is the lapse rate describes the probability of a response independent of the stimulus magnitude (e.g. lapse in concentration).

The method of fitting the PF was implemented using the Palamedes toolbox (Prins and Kingdom, 2018). Once the estimation of the PF parameters was complete, the routines for estimating the errors (standard error associated with the location and slope estimates) and the goodness of fit (deviation and associated *p*-value[7]) were carried out. The Palamedes toolbox estimates the errors and goodness of fit by simulating sets of data based on the actual data, and estimating the PF parameters repeatedly. The number of simulated sets of data generated was 400, as recommended by Prins and Kingdom (2016). The lapse rate, or the rate at which errors occur due to lapses in attention, independent of stimulus characteristics, was set to 1% (0.01). This was done to avoid bias in the fitting process, as recommended by Klein (2001), Wichmann and Hill (2001) and Prins and Kingdom (2016). The guess rate was set to 0%, as determined by the nature of the task. The lapse and guess rate were fixed parameters, whereas the location and slope were set as free parameters, meaning the algorithm searches for the best-fitting values.

This experiment was approved by the University of Southampton Ethics and Research Governance board (Ethics number = 23783).

---

[6] Should one be concerned with representing the slope of the PF in units linked to the experimental task (e.g. percent correct per dB), a conversion must be performed. The reader is directed to Strasburger (2001) for a full description of how to convert between psychometric functions slope estimates. For the reader of this thesis, to convert between the estimated β parameter (of the logistic function) and the equivalent p(yes)/dB value, the β value should be multiplied by 4.

[7] For full details of the deviation and associated *p*-value, see Prins and Kingdom (2016). The *p*-value, sometimes called 'pdev', is the key value for assessing a PF's goodness-of-fit and will have a value between 0 and 1. The higher the value, the better the fit. By arbitrary convention, a PF fit is considered unacceptably poor if the pdev value is less than 0.05.

5.4 Results part 1: Initial analysis and fitting of psychometric functions

5.4.1 Overview of results

First, the effect of stimulus level on the proportion of yes responses (p(yes)) for individual conditions was explored. Then the fitting of psychometric functions (PFs) is reported. Some conditions had observations that led to problematic PF fits, so an additional procedure to the planned analysis was used to reduce the effects of the problematic fits. This procedure is described in full in this section. The next section (Section 5.5) analyses the psychometric functions.

It was considered important that the reader can examine individual participant responses and psychometric functions. Due to the large number of conditions and participants, individual functions and descriptions of the psychometric functions are shown in Appendix E, in section E.2. These figures display the raw data and the fitted psychometric functions.

5.4.2 Effect of stimulus level

It is clear from examining the figures in Section E.2 that for the vast majority of conditions and participants, the p(yes) responses transition from p(yes) = 0 at a relatively low level, to p(yes) = 1 at a higher level (see Figure 5.3 for examples). This indicates:

1. When the subject is presented with a stimulus with a relatively low level, their judgement is that the target would not be able to detect that sound.
2. When the subject is presented with a stimulus with a relatively high level, their judgement is that the target would always be able to detect that sound.
3. The transition between these two observations occurs over a range of intensities, such that at some levels, the subject judges the target to be able to detect the sound on some trials, but not all.

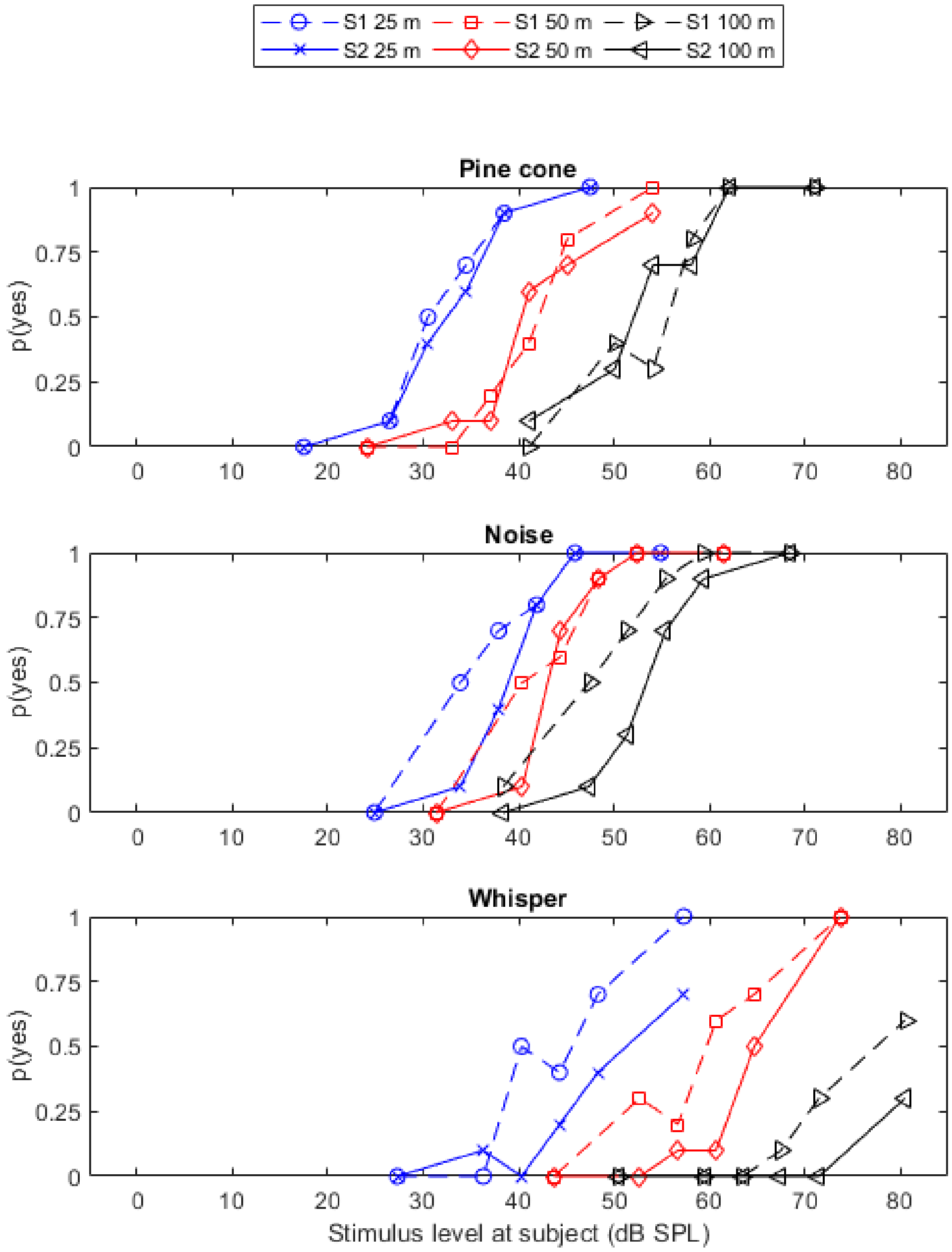

Figure 5.3 - Individual results for participant 1. Each panel represents the results for one stimulus (pine cone, noise and whisper). Each line represents one target distance (blue = 25 m, red = 50 m, black = 100m). Dashed and solid lines represent sessions 1 and 2. The stimulus level is the root-mean-square sound pressure level (in dB) at the subject's ear. P(yes) shows the proportion of 'yes' responses at each stimulus level.

These observations are in line with expectations about how participants would behave with regard to judgements of 'no' at low levels and 'yes' at high levels. However, upon close inspection of the figures in section E.2 (Appendix E), there are some conditions that do not show this relationship. For some participants, there are some conditions where the highest p(yes) is less than 1. This means that even at the highest stimulus level they did not always respond 'yes'. Likewise, there are some conditions, for some participants, where they did not always respond ‘no’ at the lowest stimulus level. An example of this is given in Figure 5.3, which shows the results for an individual participant. The majority of results shown in Figure 5.3 satisfy statements 1 to 3 given above; however, in the whisper 100 m (session 1 and 2), the pine cone 50 m (session 2) and the whisper 25 m (session 2) conditions, there is not a stimulus level that results in a p(yes) = 1.

For these conditions (with a maximum p(yes) < 1), it appears that the participant was not presented with stimuli at levels high enough to satisfy their criterion for responding ‘yes’ all the time. In other words, the method used in this experiment, for some conditions, has led to an incomplete function. The PFs in Appendix E also show a small number of conditions with incomplete functions where p(yes) = 0 was not measured. It is most likely that these conditions also involved presenting the subject with a range of intensities that did not allow the transition between p(yes) = 0 and p(yes) = 1 to be measured.

There are two likely factors that resulted in incomplete functions being measured in some conditions. Firstly, the range of levels used in the experiment were informed by pilot studies. More extensive pilot work may have identified the need for higher maximum stimulus levels, thereby increasing the likelihood of measuring full functions for all participants. Secondly, the familiarisation period of the experiment was used to identify the range of levels to use during the main data collection period, on an individual basis. For the majority of participants and conditions, this method seems to have been appropriate; however, some conditions would have benefitted from a different range of levels. Given that the familiarisation period and subsequent stimulus level range decisions were based on just 2 repeats at a variety of levels at the start of the session, it is reasonable to expect some erroneous judgements to be made that do not reflect the true function.

Future experimental work on aural detectability judgements that uses the method of constants should consider a different method for identifying which stimulus levels to use, should the method of constants be the psychophysical method. Phatak et al. (2018) used the method of adjustment to identify the speech reception threshold signal-to-noise ratio whilst measuring many psychometric functions using a speech identification task; the authors commented that this novel approach worked very effectively in terms of accuracy and time efficiency. This approach could work effectively for aural detectability judgements.

In 14% of cases, p(yes) did not grow monotonically with stimulus level (e.g. Figure 5.3, bottom panel, farthest left curve). The lack of consistency across sessions and participants suggests this is likely a result of measurement error (due to the low number of trials per stimulus level) rather than true non-monotonicity.

In summary, the effect of stimulus level on the proportion of yes responses for a given condition was largely as expected. The results show:

- When participants were presented with stimuli at relatively low levels, they generally judged those stimuli to be undetectable to the target. When participants were presented with stimuli at relatively high levels, they generally judged those stimuli to be detectable all of the time to the target.
- A monotonic relationship usually existed where the proportion of 'yes' responses transitioned between 0 and 1.
- Some conditions resulted in an incomplete measurement of the PF, most likely due to using a set of stimulus levels that did not permit full measurement of the function.

It can be concluded that aural detectability judgements are sensitive to the level of the stimulus in any one condition.

5.4.3 Psychometric functions

The next stage of analysis involved characterising the results by fitting a psychometric function (PF) to the data for each condition. As discussed in the previous section, some participants had conditions with incomplete functions due to an inappropriate set of stimulus levels. A strategy for assessing the fitted PFs and dealing with problematic data was created in order to move forward with the analysis. The three-stage strategy is first summarised as the following, and then described and justified.

1. Assess the fit of the psychometric function.
2. Distributions of PF locations and slopes assessed for outliers and normality.
3. Missing data imputation.

5.4.3.1 *Assessing the fit of the functions*

Following the fitting of PFs, each individual PF was assessed to determine whether each individual PF described the results adequately or appropriately, and subsequently determine whether or not to use the data in later analysis. This was necessary due to the reasons previously discussed: some conditions resulted in incomplete functions being measured; also, it was apparent in the figures in Appendix E.2 that some conditions resulted in poor PF fits that do not accurately represent the data. The following strategy for determining the quality of fits and how to move forward with the analysis was used. It was decided that 4 categories could adequately describe the fits:

- Acceptable
- Tolerable
- Incomplete
- Unacceptable.

To determine each PF's category, the fit was first examined visually to determine whether sufficient p(yes) measurements had been taken to permit accurate estimation of the location and slope; then the standard errors (SE) of the location and slope and the deviation and associated *p*-value were examined.

It was decided that an 'acceptable' fit depends on several criteria. Firstly, at least 1 p(yes) value between the upper asymptote and p(yes) = 0.5 point, and at least 1 p(yes) value between the lower asymptote and p(yes) = 0.5 point must have been measured. This ensures that the PF estimation process includes observations that are placed on the transitions between the upper and lower asymptotes. To ensure that the PF was based on multiple p(yes) measurements where the subject was at least sometimes judging 'yes', a criterion of at least 3 p(yes) measurements above p(yes) = 0 was included. The location and slope had to have maximum SEs of 5 dB and 2, respectively. The location SE criterion was decided by considering the differences we are interested in discriminating between conditions when investigating effects of distance and stimulus type, and setting the acceptable error accordingly. The PF locations in this dataset ranged from 18 dB SPL, up to 90 dB SPL so a very small SE criterion (e.g. 1 dB) would have been too conservative. Based on the observed range of locations, a 5 dB SE of the location was considered an appropriate limit. Finally, a deviation p-value of greater than 0.1 was considered to be an acceptable fit. If a condition's PF met all of these criteria, the location and slope were not removed in the data set.

A 'tolerable' fit had at least 1 p(yes) value at or above 0.5, and at least 1 p(yes) value at or below 0.5. These criteria were chosen to enable the identification of PFs that we based on limited but appropriate data. PF fits under these circumstances are not ideal because the location and slope estimates are likely to have a greater degree of error associated with them than the 'acceptable' fits. However, with one high p(yes) value (e.g. above 0.5) and one low p(yes) value (e.g. below 0.5), the location of the PF probably lies somewhere between these two values. It is therefore wasteful to simply reject these data. However, given the limited data points, it is unlikely that a good estimate of the slope can be obtained, so in these conditions, the slope were removed from the data set. Provided that the estimate of the error of the location is tolerable, the data still adds a valuable contribution for investigating group effects. For this reason, a criterion of a maximum location SE of 5 dB was assigned. A final criterion for a 'tolerable' fit was a deviation p-value greater than 0.05. If a condition's PF met all of these criteria, the location remained in the data set.

An 'incomplete' fit occurred when all p(yes) values were above or below p(yes) = 0.5. This criterion captures those conditions where either an inappropriate set of stimulus intensities was presented. In conditions where the participant was presented with the highest or lowest possible stimulus intensity, the fact they did not respond as expected represents an interesting finding. Removing these data from the dataset would not capture what these conditions represent. That is, for these conditions, the location of the function was particularly high or low, but could not be measured due to methodological limitations. A solution to this predicament is to allocate a value for the location that is slightly above the highest (or lowest) stimulus intensity. This preserves what the participant's responses reflect within subsequent analyses with the caveat that the true location value is *at least* this value. This is similar to how pure tone thresholds are measured when the limit of an audiometer or transducer is reached, according to recommended clinical procedures (BSA, 2011). Conditions where an incomplete function was measured and the maximum or minimum set of stimulus levels was presented had their locations recorded as 5 dB less or more than the minimum or maximum presentation level for the location, respectively. For example, is a participant was presented with stimuli at 50.5, 59.5, 63.5, 67.5, 71.5 and 80.5 dB SPL (the maximum set of stimulus levels available) and if the p(yes) at 80.5 dB SPL was less than 0.5, the location was replaced as 85.5 dB SPL. The slopes for these conditions were recorded as missing. Conditions where an incomplete function was measured and the participant was not presented with the maximum or minimum set of stimulus levels had their location and slope data recorded as missing.

A PF fit was considered 'unacceptable' if the SE of the location or slope was over 5 dB or 2 respectively. In line with the recommendations set out in Prins and Kingdom (2016), the somewhat arbitrary criterion of 0.05 was used as the maximum permissible deviation p-value. If the PF had a deviation p-value of less than 0.05, the PF was considered to not give an adequate fit to the data and therefore considered 'unacceptable'. Conditions meeting these criteria were removed from the dataset.

Using the criteria detailed above, each PF was examined, labelled and actioned as one of the following according to the respective criteria, summarised here:

- *Acceptable*.
  - There were at least 3 p(yes) measurements above p(yes) = 0
  - There was at least 1 p(yes) measurement between p(yes) = 0.1 and p(yes) = 0.5
  - There was 1 p(yes) measurement between p(yes) = 0.4 and p(yes) = 0.9
  - The SE of the location and slope was less than 5 dB and 2, respectively
  - The p-value associated with the deviation was greater than 0.1
  - The location and slope of *acceptable* conditions were not removed from the dataset.
- *Tolerable.*
  - There was at least 1 p(yes) measurement at or above p(yes) = 0.5.
  - There was at least 1 p(yes) measurement at or below p(yes) = 0.5
  - The SE of the location and slope was less than 5 dB and 2, respectively.
  - The p-value associated with the deviation was greater than 0.05.
  - The location of *tolerable* conditions was not removed from the dataset, but the slope was removed.
- *Incomplete*
  - There were no p(yes) measurements above p(yes) = 0.5, or there were no measurements below p(yes) = 0.5.
  - If the participant was presented with the highest range of stimulus intensities, the location was recorded at 5 dB greater than the maximum stimulus level. The slope was recorded as missing data.
  - If the participant was presented with the lowest range of stimulus intensities, the location was recorded as 5 dB less than the lowest stimulus level. The slope was recorded as missing data.
  - If the participant met the first 'incomplete' criterion but was not presented with the highest or lowest range of stimulus intensities, the location and slope was recorded as missing.
- *Unacceptable*
  - The SE of the location and slope was greater than 5 dB and 2 dB, respectively.
  - The p-value associated with the deviation was less than 0.05.
  - The location and slope of *unacceptable* conditions were removed from the dataset and considered missing data.

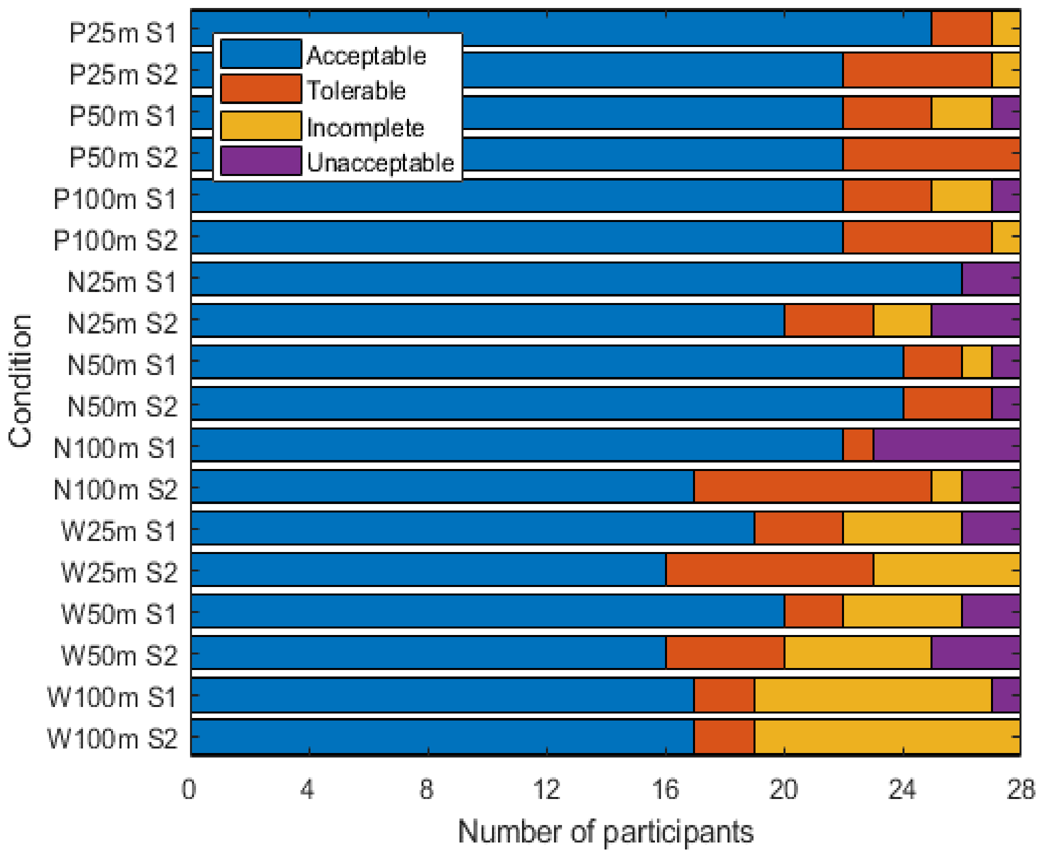

Figure 5.4 - Breakdown of psychometric function fit categories for each condition. Conditions are represented by the following: P = pine cone, N = noise, W = whisper; target distance 25, 50 or 100 m; S1 and S2 is session 1 and 2 respectively.

Figure 5.4 shows a breakdown of the PF fit categories for each of the 18 conditions (2 sessions x 3 distances x 3 stimulus types). Each bar shows how many of the 28 participants had acceptable, tolerable, incomplete and unacceptable PF fits.

The majority of PF fits met the 'acceptable' and 'tolerable' criteria. The number of 'tolerable' and 'unacceptable' fits was roughly independent of stimulus type, distance and session. There was a larger number of 'incomplete' fits for the whisper stimuli than for the noise and pine cone stimuli (35 out of 46 'incomplete' fits were whispers). Furthermore, of the 35 'incomplete' whisper functions, 26 were based on the highest set of stimulus intensities. None of the 46 'incomplete' fits resulted from the lowest set of stimulus intensities; all were either an inappropriate (19 out of 46) or the highest possible (27 out of 46) set of stimulus levels.

Whilst the method for identifying which set of stimulus levels to use for each individual has been highlighted previously as a weakness, the number of incomplete functions due to an inappropriate set of levels shows that this weakness had a moderate (19 out of 504 PFs had inappropriate levels).

5.4.3.2 *Identifying outliers and assessing distributions*

The next stage involved identifying outliers and assessing the distribution of the locations and slopes from the remaining data to determine whether the planned parametric statistical analyses could be conducted. The removal of outliers was not planned; it became apparent that an outlier removal strategy was required following the fitting of psychometric functions. First, the location and slope data were converted to z-scores and any data point with a z-score greater than 3, or less than -3, were excluded. Of the 504 PFs (28 participants x 18 conditions), 44 locations were missing due to unacceptable or incomplete PFs and 1 location met the z-score exclusion criteria, leaving a total number of locations in the sample as 459. For the slopes, 131 slopes were removed due to being categorised as tolerable, unacceptable or incomplete PFs, and 8 met the z-score exclusion criteria, leaving the number of slopes in the sample as 365.

Due to the high number of removed slope data, it was decided that the data would be pooled across the two sessions. This was achieved by taking the mean slope value of session 1 and 2.. This prevented the effect of session on the slope of the PF being explored, but it was considered preferable to do this in order to have a more complete sample. Following pooling the slope values, 219 out of 252 conditions (9 conditions x 28 participants) had a slope value.

Next, the remaining data were displayed in box plots for the purpose of visually assessing the distributions for normality and skewness. Figure 5.5 shows the locations (n = 459) and Figure 5.6 shows the slopes (n = 219). The coloured boxes show the interquartile range; the horizontal black line in the box shows the median; the tails show the range; outliers are shown by coloured circles and are defined as a data point outside 1.5 x the interquartile range; extreme outliers are shown as asterisk and are defined as outside 3 x the interquartile range. The location box plots show just two outliers and broadly normal distributions for each condition. The slope box plots show 6 outliers (circles) and 1 extreme outlier (asterisks). The participant number corresponding to the outlying data point is shown. Due to the prior z-score data cleaning, it was decided that no further outlier removal would occur, apart from the extreme outlier in the Noise 100 m condition, participant 18, as this could strongly affect the statistical analysis

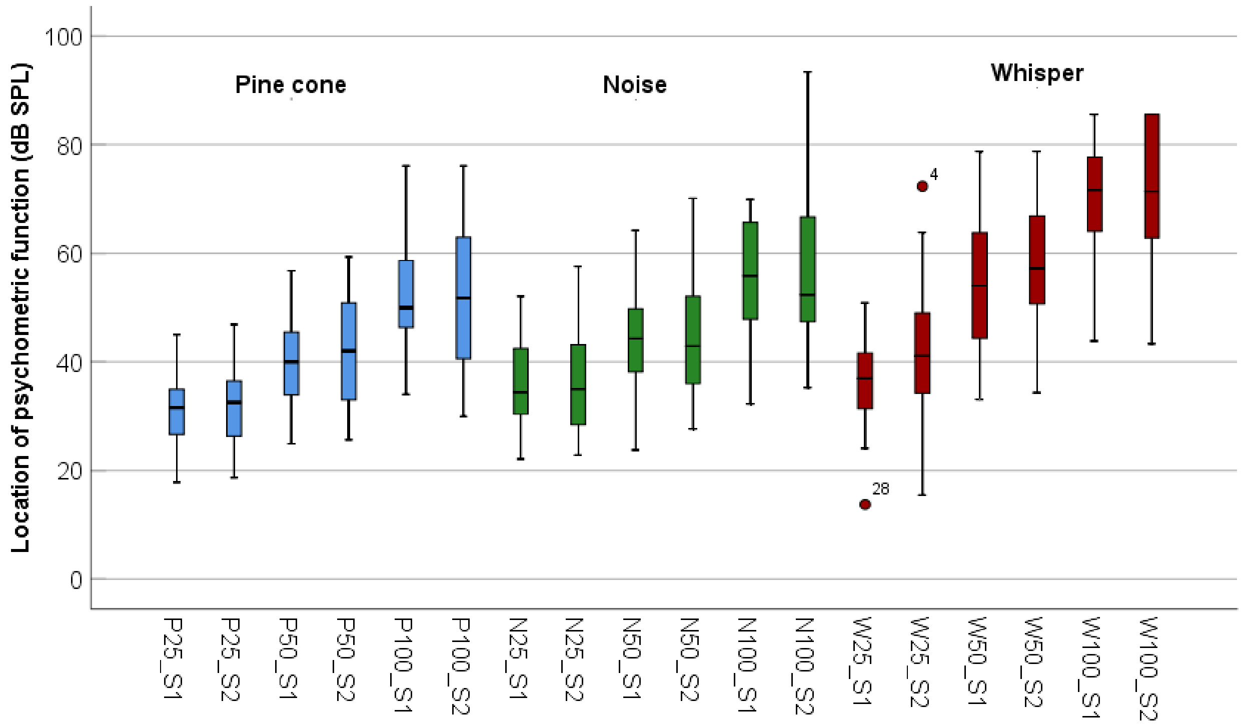

Figure 5.5 - Box plots showing the locations of psychometric functions in Experiment 2. Blue, green and red boxes represent pine cones, noise and whisper conditions, respectively. The target distance and session number are shown on the x axis. The numbers by the circles show the participant number.

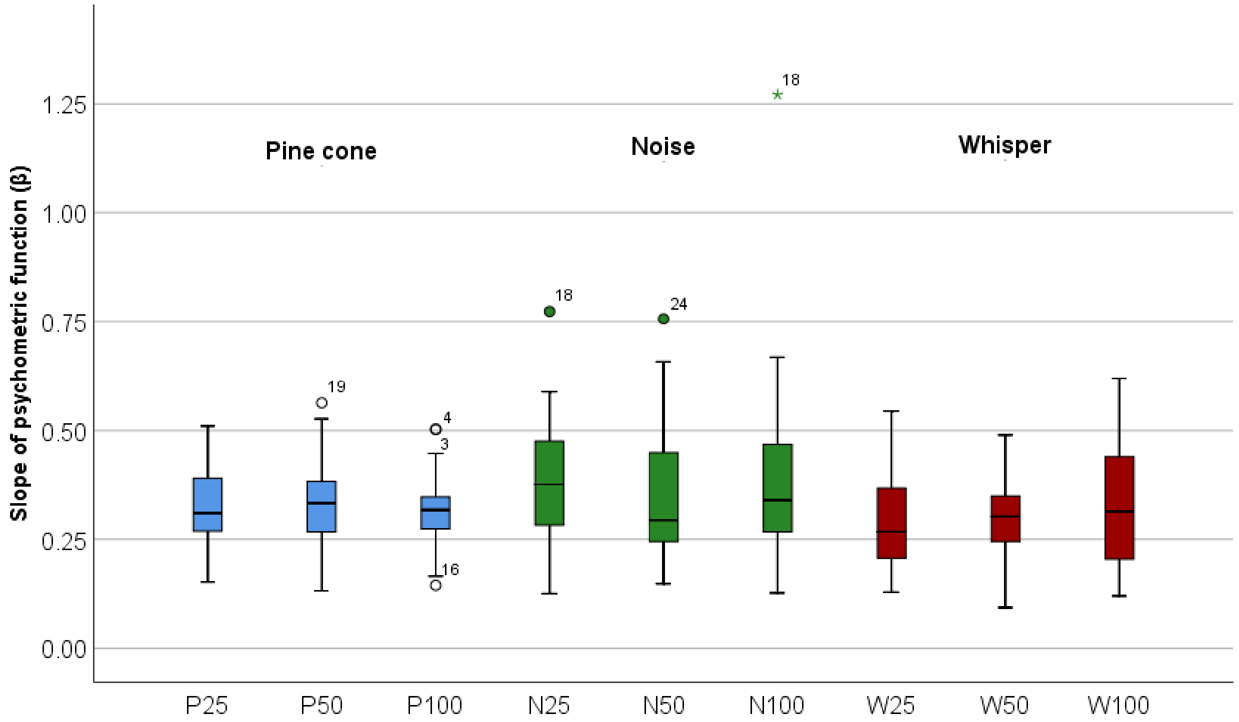

Figure 5.6 - Box plots showing the slopes of psychometric functions in Experiment 2. Blue, green and red boxes represent pine cone, noise and whisper, respectively. The target distance are shown on the x axis. The numbers by the circles and asterisk show the participant number for outliers.

The assumption of normality in each condition was tested by performing the Shapiro-Wilk test of normality (Shapiro and Wilk, 1965) and visually examining the box plots. Note, the data assessed here are those which followed the previous section's treatment (z-score removal criteria and PF category removal strategy). No conditions yielded a significant test result, suggesting no deviations from normality.

5.4.3.3 *Missing data imputation*

As a result of the PF assessment criteria and consequent removal of some data, there were missing values within the dataset (8% of locations, 14% of slopes). Different strategies could be implemented to deal with the missing data, each with different merit (Baraldi and Enders, 2010). Two common approaches are case-based deletion and imputation (Peugh and Enders, 2004). Case-based deletion excludes a subject's data from the statistical test if that subject does not have a complete set of data. This approach excludes their data from all conditions (for ANOVA), which, since the data takes many hours to collect, is not recommended (Little and Rubin, 2014). Also, for the present data set, only 13 out of 28 participants have 'complete' sets of location data, meaning that half the sample would not be included. For the slope data, no participants have a full set of data, meaning no statistical analysis would be possible. This approach is too conservative on this dataset and was therefore not adopted.

The imputation approach effectively replaces missing data with values that are derived based on the rest of the sample. Different approaches can be used to calculate the imputed value. The approach used can depend on the reason for the missing data and the type of data (Rubin, 1976, Baraldi and Enders, 2010). The simplest approach is to replace the missing value with the mean of the remaining data for that condition, known as mean imputation. This approach does not substantially affect the mean of the sample but can lead to biases in the measures of variation (Peugh and Enders, 2004). More advanced techniques, such as Maximum Likelihood Estimation or Multiple Imputation, create simulated datasets and produce imputed values that include some simulated variability (Schafer and Graham, 2002). For simplicity and ease of implementation, mean imputation was used for the present dataset.

For the locations, all missing values (45 out of 504) were replaced by the mean value for that condition. For the slopes, the missing data following the pooling of sessions 1 and 2 (33 out of 252) were replaced with the mean for that condition.

5.4.4 Summary of part 1 results

The first part of the results of Experiment 2 showed that:

- Aural detectability judgements were sensitive to stimulus level in all conditions, where the subject tended to report the sound as inaudible to the target at relatively low levels and audible to the target at higher stimulus levels.
- In a small number of conditions, incomplete functions were measured, suggesting that the range of stimulus levels was not appropriate in those conditions.
- The majority of psychometric function fits were acceptable, but some were poor due to the aforementioned methodological limitations. A *post hoc* strategy was employed to identify poor fits and outliers; these data were removed and the missing values were replaced by imputation.

5.5 Results part 2: Analysis of psychometric functions

5.5.1 Exploration of average psychometric functions

The average PF location and slope for each condition are shown in Table 5-2. The mean and standard deviation were calculated for each condition from the dataset following outlier removal and missing data imputation. The location of PF corresponds to the stimulus level that subject judged the sound to be audible to the target 50% of the time. The slope of the PF shows the steepness of the function, with a higher value indicating a steeper slope. Note, the stimulus level for the location of PF is expressed in terms of the RMS level that the subject experienced. The mean psychometric function location for each condition is shown in Figure 5.7. The mean slope for each condition is shown in Figure 5.8.

Table 5-2 - Mean and standard deviations psychometric function location and slopes for each condition in Experiment 2.

Stimuli	**Target distance (m)**	**Session**	**Location (dB SPL)**		**Slope value**	
			Mean	*Standard deviation*	*Mean*	*Standard deviation*
Pine cone	25	1	30.8	6.8	0.33	0.08
		2	32.0	7.2		
	50	1	39.9	8.3	0.33	0.10
		2	41.4	10.2		
	100	1	51.9	10.6	0.32	0.08
		2	52.1	12.2		
Noise	25	1	35.5	7.7	0.39	0.14
		2	36.5	9.0		
	50	1	44.5	9.6	0.34	0.15
		2	45.2	11.6		
	100	1	55.9	9.7	0.37	0.14
		2	56.3	13.0		
Whisper	25	1	36.2	7.3	0.29	0.10
		2	41.8	11.7		
	50	1	55.9	13.2	0.29	0.09
		2	59.4	11.3		
	100	1	70.2	11.6	0.33	0.13
		2	70.8	13.0		

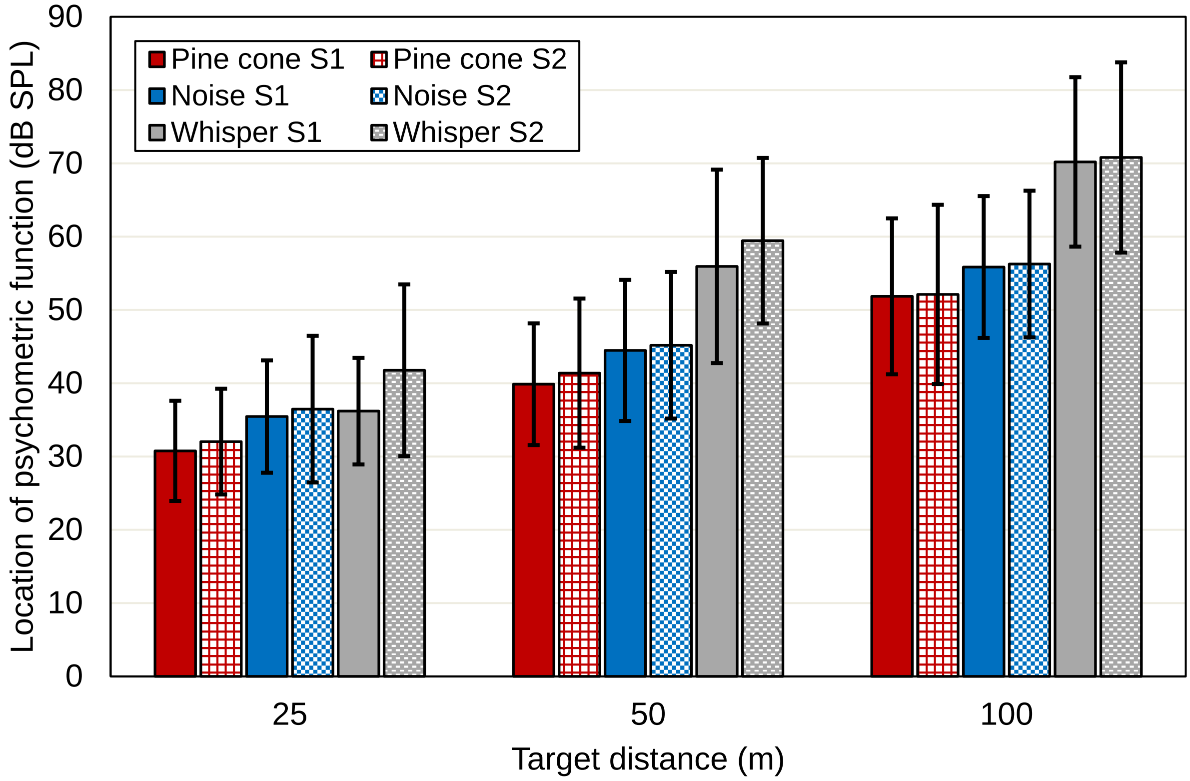

Figure 5.7 - Mean location of psychometric function in Experiment 2. The red bars are for the pine cones, blue bars for the noise and grey bars for the whisper. The solid coloured and chequered bars show the locations for session 1 and 2, respectively. Error bars show ±1 standard deviation.

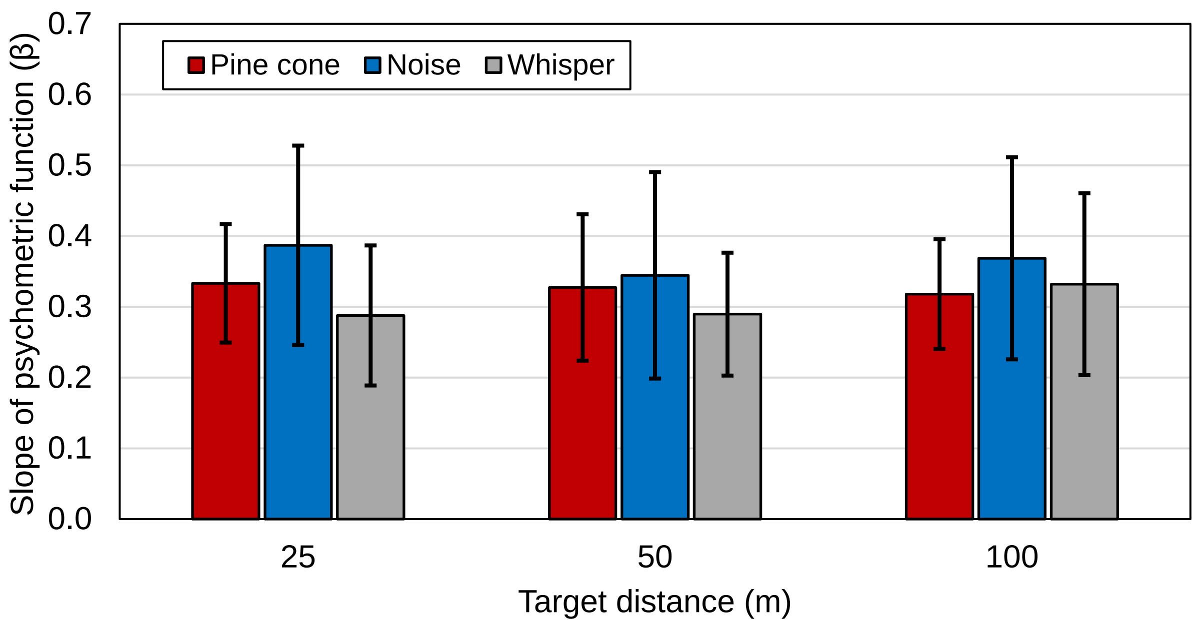

Figure 5.8 - Mean slopes of psychometric functions in Experiment 2. Red, blue and grey bars are for the pine cone, noise and whisper stimuli, respectively. Error bars show ±1 standard deviation.

The mean locations of the PFs increase as the target distance increases. This trend is apparent for all three stimulus types. The effect of distance appears to be greatest for the whisper stimuli, with the greatest difference between target distances (as seen in Table 5-2).

There appears to be an effect of stimulus type on the location of psychometric function, with pine cones having the lowest mean location, then noise, then whisper. This ranking of stimulus type is consistent at each target distance. The difference between whispers and the other two stimuli appears to be greater at 50 and 100 m target distance, suggesting the effect of stimulus type might depend on target distance.

The mean location of the PF appears similar in all conditions between session 1 and 2, with mean differences of approximately < 2 dB, with the exception of the whisper 25 m condition (5.6 dB). The average difference between session 1 and 2, albeit small, is consistently in the direction of session 2 being higher than session 1. This suggests that there might be a small systematic difference between session 1 and 2.

There does not appear to be an effect of target distance on the slope of the PF. The rank order for each stimulus type is the same for each target distance, suggesting that there might be an effect of stimulus type on the slope of the PF.

5.5.2 Statistical analysis of psychometric function location and slope

5.5.2.1 *Description of ANOVAs*

Statistical analyses further examined the locations and slopes of PFs that were explored in Section 5.5.1.For the locations, expressed as the RMS sound pressure level at the subject, a repeated-measures ANOVA was conducted with factors: Stimulus (Pine cone, noise, whisper), Target distance (25, 50, 100 m) and Session (session 1 and 2). Mauchly's test showed that the assumption of sphericity was violated for the main effects of stimuli, $\chi^2(2) = 13.5$, $p < 0.05$, and target distance, $\chi^2(2) = 12.2$, $p < 0.05$, and the interactions between stimuli and target distance, $\chi^2(9) = 42.6$, $p < 0.05$, and stimuli and session, $\chi^2(2) = 10.4$, $p < 0.05$, and the interaction between distance, stimulus and session, $\chi^2(9) = 17.6$, $p < 0.05$. Therefore, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity. The F-ratios for this ANOVA are shown in Table 5-3.

For the slopes, a two-way repeated-measures ANOVA was conducted with factors: Stimulus (Pine cone, Noise, Whisper) and Target distance (25, 50, 100 m). Note, the test was carried out on the pooled slopes (across sessions 1 and 2) as discussed in Section 5.4.3. Mauchly's test revealed that the assumption of sphericity was not violated for either main effect or the interaction between stimuli and distance. The F-ratios for this ANOVA are shown in Table 5-4.

Table 5-3 - F-ratios for the locations of the psychometric functions, expressed as the sound pressure level at the subject, in Experiment 2. df = degrees of freedom. Significant results are highlighted in yellow.

Source	***df* (source)**	***df* (error)**	***F*-ratio**	**Sig.**
Stimulus	1.42	38.45	55.91	< 0.001
Distance	1.46	39.31	387.20	< 0.001
Session	1.00	27.00	3.49	0.073
Stimulus * Distance	2.21	59.78	25.91	< 0.001
Stimulus * Session	1.51	40.63	3.26	0.062
Distance * Session	2.00	54.00	2.18	0.123
Stimulus * Distance * Session	2.97	80.12	1.45	0.234

Table 5-4 - F-ratios for the slopes of psychometric functions in experiment 2. df = degrees of freedom. Significant results are highlighted in yellow.

Source	***df* (source)**	***df* (error)**	***F*-ratio**	**Sig.**
Stimulus	2	54	9.25	< 0.001
Distance	2	54	0.76	0.471
Stimulus * Distance	4	108	1.47	0.215

5.5.2.2 *Test-retest reliability*

The test-retest reliability of aural detectability judgements was analysed using the repeated-measures ANOVA, correlation of results between sessions 1 and 2, and by calculating the 95% confidence associated with any one measurement, as discussed in Section 2.4.5.4. Only the test-retest reliability of the location of the PF could be analysed due to the pooling of slopes.

The repeated-measures ANOVA on the locations of the PFs revealed no significant main effect of session or interaction involving session. The main effect of session and the interaction between session and stimulus type nearly reached significance, suggesting that a trend may be present. The grand mean difference between the locations of the PF for sessions 1 and 2 was 1.7 dB (session 2 higher than session 1; 95% confidence interval of the difference = -0.2 to 3.5 dB). The absence of a statistically robust interaction suggests that the lack of effect of session is consistent in all conditions.

The replicability of location estimates was measured by calculating the Pearson's correlation coefficient, shown in Table 5-5. All correlations were in the 0.7 – 1.0 range, suggesting strong replicability.

Table 5-5 - Correlations between locations of psychometric functions between session 1 and 2.

Stimuli	Distance (m)	Pearson's correlation coefficient, r	Sig.
Pine cone	25	0.836	< 0.0001
	50	0.858	< 0.0001
	100	0.870	< 0.0001
Noise	25	0.903	< 0.0001
	50	0.884	< 0.0001
	100	0.918	< 0.0001
Whisper	25	0.715	< 0.0001
	50	0.927	< 0.0001
	100	0.886	< 0.0001

The 95% confidence interval associated with an individual measurement (i.e. the range of values the test would give if repeated many times, with probability 0.95) is shown for each condition, in Table 5-6. The calculation of these statistics is described in section 2.4.5.4. The mean difference between session 1 and 2 is also included in the table as a measure of stability.

Table 5-6 - Statistics associated with the test-retest reliability for the measurement of the location of psychometric functions in Experiment 2. For each condition, the 95% confidence interval associated with a single measurement and the mean difference between sessions is shown.

Stimulus	Target distance (m)	95% confidence interval associated with a single measurement (dB)	Mean difference between session 1 and 2 (dB)
Pine cone	25	5.6	1.3
	50	8.9	1.5
	100	10.2	0.3
Noise	25	6.8	1.0
	50	10.5	0.7
	100	12.8	0.4
Whisper	25	12.3	5.6
	50	13.4	3.5
	100	10.4	0.6

These measures of test-retest reliability show that there was no statistically robust difference between locations of the PFs between sessions. This suggests that subjects do not systematically change their judgements on different sessions and that the variability observed is largely due to random factors.

5.5.2.3 *Effect of target distance*

The statistically significant main effect of target distance ($p < 0.001$) indicated that, averaged across the other factors, the PF location changed with target distance. The significant interaction between stimulus type and target distance ($p < 0.001$) indicated that the effect of target distance depended on the stimulus type (interaction explored later in Section 5.5.2.5). Figure 5.9 shows the mean PF location across sessions 1 and 2, for each stimulus type at each target distance. As the target distance increased, the PF location increased. Post-hoc Bonferroni corrected t-tests confirmed that the difference between target distances (e.g. the difference between PF location for pine cone 25 m and pine cone 50 m), for each stimulus type, was statistically significant, indicating that the effect of target distance was present between 25 vs. 50 m, 50 vs. 100 m, and 25 vs. 100 m.

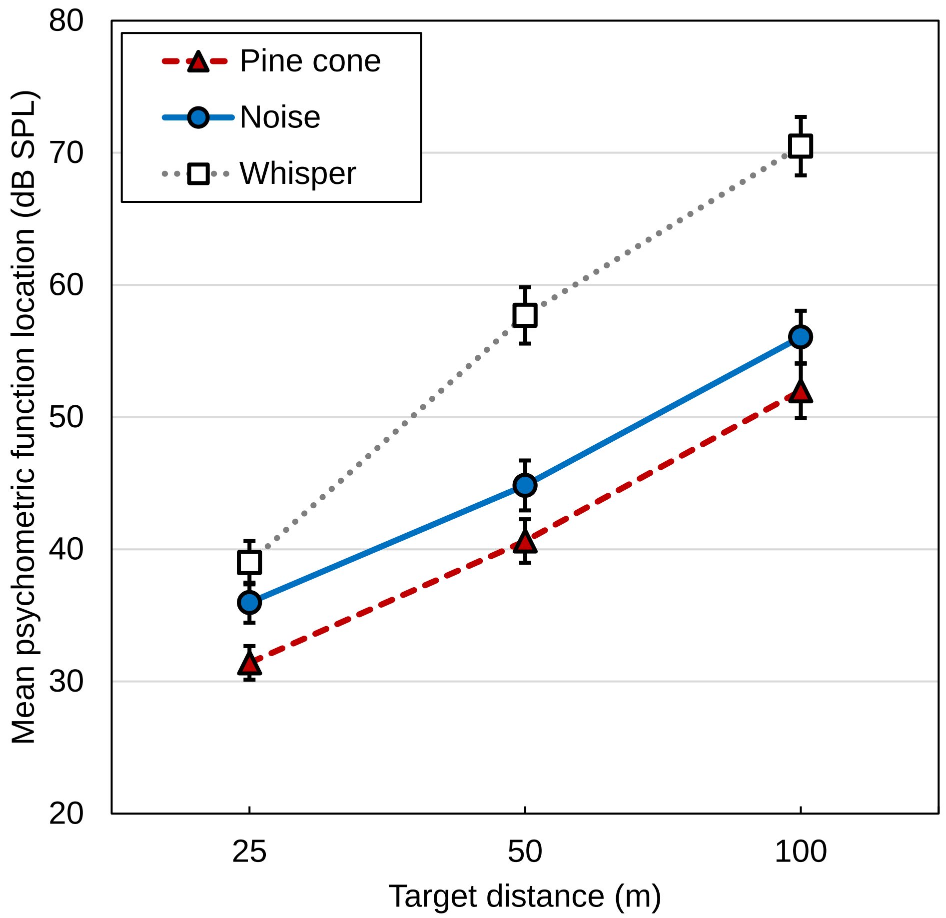

Figure 5.9 - Mean location of psychometric function for each target distance and stimulus type. Sound level represents the level at the subject location. Each point represents the mean location across sessions 1 and 2, for each condition. Error bars indicate ±1 SE.

No effect of target distance on the PF slope was observed. As discussed previously, it is not clear what precise aspect of the task the slope of the PF represents, so it is not possible to conclude what this null result represents. It is only possible to conclude that these data do not find an effect of target distance on the slope of the PF.

5.5.2.4 *Effect of stimulus type*

The significant effect of stimulus type ($p < 0.001$) indicated that, averaged across other factors, the PF location depended on the stimulus type. The interaction between stimulus type and session did not reach significance, indicating that the effect of stimulus type was consistent across sessions. Figure 5.9 shows that the whisper and pine cones led to the highest and lowest PF locations, respectively, at all target distances. This was confirmed statistically by significant post-hoc Bonferroni corrected t-tests. The interpretation of these findings is difficult due to spectro-temporal differences between stimuli; however, it can be cautiously concluded that the RMS level of the stimulus alone does not predict the PF location. This is discussed in more depth in the discussion section.

For the slopes of the PFs, a significant main effect of stimulus type was found when averaged across distances. Pairwise comparisons revealed that the noise led to a significantly different slope than the whisper and pine cone, but the slope for the pine cone was not significantly different to that for the whisper. These results are shown in Figure 5.10.

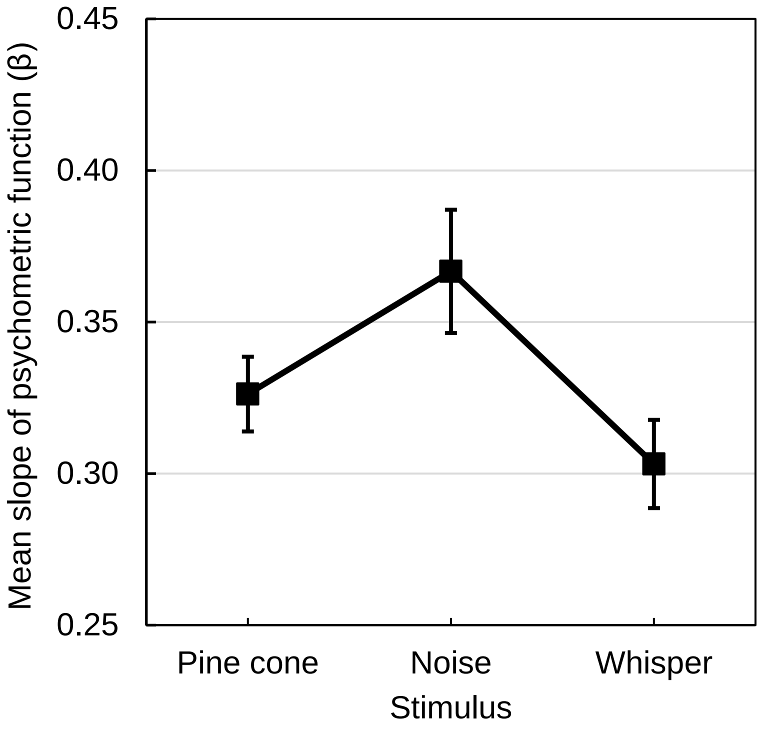

Figure 5.10 - Mean slopes of psychometric functions in Experiment 2. Errors bars indicate ±1 SE.

5.5.2.5 *Interaction between stimulus type and target distance*

The effect of target distance of the PF location depended on the stimulus type, as indicated by the significant interaction between stimulus type and target distance ($p < 0.001$). The three-way interaction (stimulus, target distance, session) did not reach significance, indicating that the interaction between stimulus and target distance was consistent across sessions. Post-hoc Bonferroni corrected t-tests were carried out to further explore the interaction. First, for each participant, the difference between the location of the PF at 25 and 50 m, 25 and 100 m, and 50 and 100 m, for each stimulus, was calculated. Then, a 2-tailed paired-sample t-test was performed comparing the difference in PF locations between target distances for each pair of stimuli. A significant t-test shows that the difference in PF location between target distances differed across stimuli. Table 5-7 shows the results of these t-tests.

Table 5-7 - Post-hoc paired-sample t-tests, comparing the difference in PF location between each target distance for each pair of stimuli. Bonferroni corrected significance level = 0.0056 (SD = standard deviation; SEM = standard error of the mean; CI = confidence interval). Significant results are highlighted in yellow.

Distance comparison	**Stimulus comparison**	**Mean difference**	**SD**	**SEM**	**95% CI of difference**		**t**	**df**	**Sig.**
					Lower	*Upper*			
25 vs 50 m	PC vs N	0.35	4.33	0.82	-1.33	2.03	0.43	27	0.67
	PC vs W	-9.49	6.68	1.26	-12.08	-6.90	-7.52	27	<0.001
	N vs W	-9.84	8.44	1.60	-13.12	-6.57	-6.17	27	<0.001
25 vs 100 m	PC vs N	0.48	5.80	1.10	-1.76	2.73	0.44	27	0.66
	PC vs W	-10.93	9.39	1.77	-14.57	-7.29	-6.16	27	<0.001
	N vs W	-11.41	10.92	2.06	-15.65	-7.18	-5.53	27	<0.001
50 vs 100 m	PC vs N	0.13	3.93	0.74	-1.39	1.66	0.18	27	0.85
	PC vs W	-1.44	6.49	1.23	-3.95	1.08	-1.17	27	0.25
	N vs W	-1.57	6.65	1.26	-4.15	1.01	-1.25	27	0.22

The significant results indicate that the effect of target distance was different for the whisper stimulus compared to the noise and pine cone. The effect can be seen in Figure 5.9. The difference between 25 and 50 m, and 25 and 100 m, was higher for the whisper stimuli than for the noise and pine cones. For the whisper stimuli, subjects generally judged the whispers to become less audible to the target at a higher rate as the target distance increased, compared to the noise and pine cone. This suggests that subjects' judgements about how sound attenuates over distance were dependent on the stimulus.

5.5.3 Egocentric distance estimates

The median and interquartile ranges of visual distance estimates are shown in Figure 5.11. A repeated-measures ANOVA, with target distance (25, 50 and 100 m) and session (session 1 and 2) as factors, showed a significant main effect of true distance, $F(1.079,28.049) = 90.925$, $p < 0.001$ (Greenhouse-Geisser correction). The effect of session and the interaction between true distance and session did not reach significance. These results indicate that estimates were consistent, on average, between sessions, and that estimates were sensitive to target distance, where the distance estimates increased as the true distance increased.

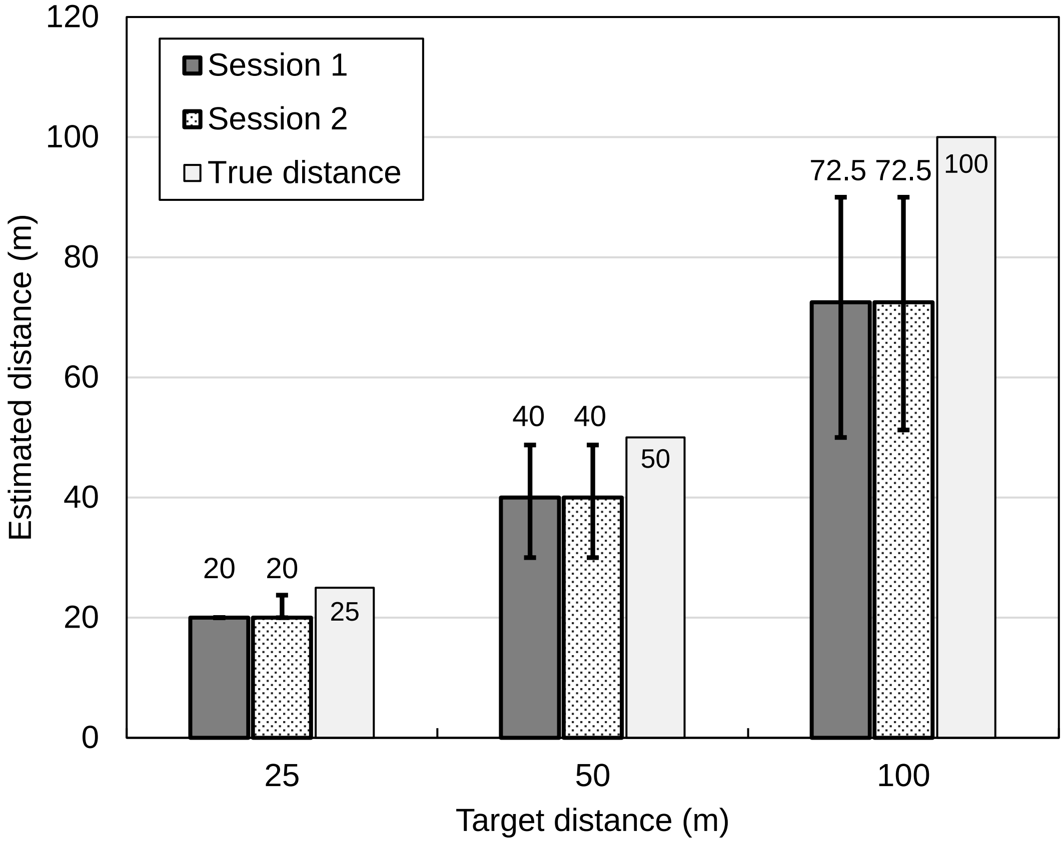

Figure 5.11 - Distance estimates in Experiment 2. Median estimates are shown by the columns and the interquartile ranges are shown by the error bars. The true distances are shown by the light grey columns.

The median estimates show that the participants, on average, underestimated the target distance by around 20%, at each distance. The ratio between judged target distances was approximately correct: the ratio of estimates between 50 to 25 m and 100 to 50m target distances is approximately 2:1. This shows that, although there is a tendency for subjects to underestimate the true target distance, relative distances were estimated well.

5.5.4 Accuracy of judgements

The accuracy of judgements was assessed by expressing the PF locations in terms of the predicted target sensation level (see Section 5.2.5 for review of justification). If a participant accurately judged aural detectability, their PF location would be 0 dB SL; this means that the mid-point of their PF was at the level equivalent to the target's threshold. A highly accurate aural detectability judgement has a PF location of close to 0 dB SL; the greater the deviation from 0 dB SL, the less accurate the subject's judgements were. A positive value indicates that the target *could* detect the sound at the subject's judged level;

a negative value indicates the target *could not* detect the sound at the subject's judged level.

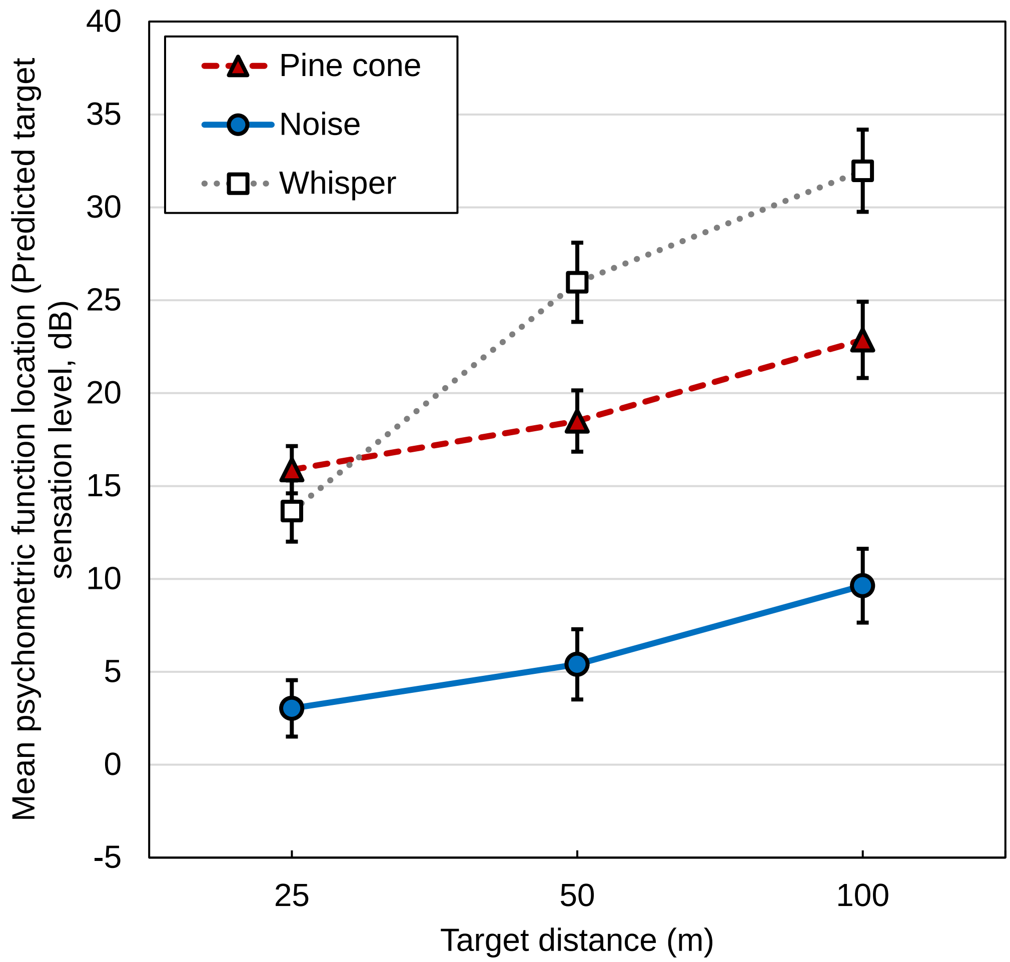

Figure 5.12 - Accuracy of aural detectability judgements in Experiment 2 according to the target sensation level. Mean psychometric functions for each stimulus type and target distance, expressed in terms of the predicted target sensation level (dB SL). Error bars indicate ±1 SE.

Figure 5.12 shows the mean location of PF, expressed as the target sensation level, as calculated by the target sensation model. The mean PF location was used due to the absence of a session effect. According to the target sensation level model, subjects, on average, gave inaccurate judgements, such that the level of the sound that they judged to be just detectable by the target was above the threshold of the target. This suggests that, on average, untrained normal-hearing civilian subjects do not give accurate aural detectability judgements.

A repeated-measures ANOVA with Stimulus and Distance as factors showed significant main effects of Stimulus, $F(1.424,38.443) = 87.335$, $p < 0.001$ (Greenhouse-Geisser correction), and of Distance, $F(1.456,39.306) = 75.778$, $p < 0.001$ (Greenhouse-Geisser correction), and a significant interaction between Stimulus and Distance, $F(2.213,59.757) = 27.250$, $p < 0.001$ (Greenhouse-Geisser correction). Pairwise comparisons with

Bonferroni corrections showed that the mean location of PF increased as distance increased for all stimuli at all levels, except for the noise stimuli between 25 m and 50 m, though this comparison very nearly reached the required significance level. The grand mean for each stimuli was lowest for noise and highest for the whisper stimuli, suggesting the whispers were least accurate. Simple contrasts revealed that the significant interaction is due to the effect of distance being significantly different for the whisper between 25 m and 50 m and between 25 m and 100 m than for the noise and pine cone. The effect of distance was not significantly different between 50 m and 100 m for any of the stimuli.

5.5.5 Summary of results

The results of experiment 2 are summarised below:

- All conditions tested showed that aural detectability judgements were sensitive to stimulus level. Subjects judged sounds to be more audible as the stimulus level increased.
- Judgements on average, and the effects of target distance and stimulus, were reasonably consistent between sessions.
- The location of the PF increased as target distance increased.
- The location of the PF depended on the stimulus type, with the whisper and pine cone having the highest and lowest average PF locations, respectively, at each target distance.
- The PF slope for noise was steeper than that for whisper and pine cone stimuli.
- The effect of target distance on the PF location depended on the stimulus type; the PF location increased more rapidly as target distance increased for whispers than for noise and pine cones.
- Visual target distance estimates were consistent between sessions, and generally underestimated the true distance by, on average, approximately 20%.
- The level judged to be just detectable to the target (the PF location) was above the target's threshold by 3 – 32 dB, according to the target sensation level model.
- The noise stimuli gave the most accurate judgements and the whispers gave the least accurate judgements, according to the target sensation level model.
  The accuracy of judgements got poorer as target distance increased for all stimuli.

5.6 Discussion

5.6.1 Overview

Experiment 2 obtained aural detectability judgements by normal-hearing civilians in a virtual environment that simulated a wide open grassy field, using a novel method. The results showed that judgements are:

- Repeatable across sessions, including the trends.
- Sensitive to stimulus level, such that (relatively) low-level stimuli are judged inaudible by the target and high-level stimuli are judged audible, and a transition occurs between these two.
- The level judged to be equivalent to the target's threshold was highest for whispers, then noise, then pine cone.
- Sensitive to target distance, where subjects judged a sound at a given level to be less detectable as target distance increased .
- The effect of target distance on judgements was different for whispers than for noise and pine cones, such that the change in PF location between 25 and 50 m was greater for whispers than for noise and pine cones.
- Judgements were inaccurate as determined by the target sensation level model. In all conditions, on average, the level subjects judged to be equivalent to the target threshold was above the target threshold, by 3 – 32 dB.
- The accuracy of judgements got worse as the target distance increased.
- The accuracy of judgements was poorest for whispers.

Sections 5.6.3 discusses the strengths and weaknesses of the methodology with a focus on how the subject experienced the experiment. Sections 5.6.4, 5.6.5 and 5.6.5 discuss the observed trends in the data. The discussion concludes by exploring the accuracy of judgements, in Section 5.6.6.

5.6.2 Methodology

Given the substantial development that occurred in order to measure aural detectability judgments, a review of the methodology used is appropriate. Overall the method worked well, though future work investigating ASA might benefit from some modifications.

The subjects tolerated testing well. They reported understanding the task rapidly and not getting bored or fatigued, partly due to the novelty of the task, and partly because of the pace. No subjects reported any nausea or symptoms of vertigo, as can sometimes occur whilst using VR headsets. Subjects sometimes reported that it took them a while to immerse themselves in the environment; some reported that they found themselves imagining background noise, as they typically associated a wide open grassy space with the sound of a park (e.g. wind noise, distant traffic, other park users). The effect of background noise on aural detectability was not included in this experiment, partly because other factors were of greater priority, and also because of the technological difficulties in doing it without measuring individual head related transfer functions. Should the effect of background noise be investigated, the use of ambisonics could be explored, as this allows effective and immersive experiences of ambient and omnidirectional noise. Alternatively, virtual acoustics could be utilised (Zahorik, 2002), though this would require significant development and specific expertise in order to integrate the virtual acoustics algorithms with the VR software.

The method of constant stimuli had limited success. Due to the large variability between subjects, this methodology resulted in a large number of incomplete PFs. It is likely that the familiarisation period was too short to gain an accurate estimate of the optimal set of stimulus levels for the main experimental trials. More familiarisation trials may remedy this unwanted effect, or a method of adjustment approach, based on the corridor pilot experiment (Section 2.4.2), could be utilised in order to estimate the location of the PF. Phatak et al. (2018) used a method of adjustment approach to estimating the SRT for a variety of conditions and then used the method of constant stimuli based on the SRT estimate; the authors described this approach as effective, so a similar method could be adopted here.

Each session took 1.5 – 1.75 hours. Whilst subjects tolerated this, it is a lot of testing to obtain a relatively small number of results. Should military personnel be tested, a reduced test time would be highly preferable. This would allow either more personnel to be tested, or more conditions to be tested. As previously discussed, accessing military personnel is difficult (Semeraro, 2016), so use of this limited time should be as effective as possible.

One way that the method could be shortened is by using an adaptive procedure rather than the method of constants. This would reduce each condition from 60 trials, to perhaps 30 or less. The potential increase in efficiency may however come at a cost to the accuracy of the data. A future experiment could compare the method of constants with an adaptive procedure in order to validate the use of an adaptive procedure. The test-retest reliability of the method was reasonable; the results showed that the trends were repeatable across session and that the score in one session was able to predict the score in the second well, though the 95% confidence interval associated with a single measurements was quite large (up to 12 dB). Should the method be adapted for efficiency purposes, the test-retest reliability should also be measured.

5.6.3 Effect of stimulus level

The psychometric functions for each condition, shown in Appendix E, illustrate that aural detectability judgements are clearly sensitive to the stimulus experienced by the subject. It was hypothesised (Section 2.3.5) that ASA was achieved by making aural detectability judgements, which comprise two components: an inference about the physical properties of the sound at the target location, and an inference about the target's perception of that sound. The physical properties of the sound at the target location, especially its level relative to a target's threshold, depend partly on the sound power of the sound source. Since the subject is close to the source, they can potentially gain reliable estimates of the sound power (assuming they have normal hearing) and infer the physical characteristics of that sound at the target. Note, the confounding effects of intensity at the ear and source distance were not present as the source distance to the subject was fixed and the subject had visual access to the source location, thereby, in theory, avoiding loudness constancy issues (Zahorik and Wightman, 2001). If subjects do factor in changing sound power, then the judgements should change as the stimulus level changes, in any particular condition. The results support this notion, which suggests that a person's ASA does involve estimating the sound power of the source. However, the effects of target distance and stimulus type, and the accuracy of judgements, suggest that, despite giving judgements that are sensitive to the sound power, sound power is not factored in correctly.

5.6.4 Effect of stimulus type

The results of Experiment 2 showed that the stimulus type affected aural detectability judgements (see Figure 5.9). The results showed that the RMS level experienced by the subject did not fully explain the judgements that subjects gave, because an effect of stimulus type was found on the location of the PF.

A limitation of this experiment is that the conditions included only allowed a limited exploration of why there was an effect of stimulus type on aural detectability judgements. This was because the experiment was primarily motivated by a practical interest in exploring judgements for sounds relevant to ASA. Given the effect of stimulus type, future work should include conditions that allow the factors that might explain the differences to be explored.

One factor that could be explored is the effect of different spectro-temporal characteristics. Because there were spectral and temporal differences between the stimulus used in the experiment (see Figure 4.5), it is possible that these were factored in to subjects' judgements. The noise stimulus had energy uniformly spread over 20 – 20000 Hz, whereas the pine cone had most of its energy in the low to mid frequencies (< 2000 Hz).

Since sound propagation and rates of attenuation are frequency dependent, subjects may have factored these differences in stimuli into their judgements. A simple way that this could be explored is by measuring judgements with filtered noise. For example, aural detectability judgements could be measured with three stimuli: an unfiltered white noise, a low-pass filtered white noise with cut-off 4 kHz and a high-pass filtered white noise with cut-off 4 kHz. Any observed differences in their judgements would be reveal how spectral differences affect judgements.

The temporal envelope differed substantially between the stimulus types included in Experiment 2. The pine cone is essentially a series of transient clicks as the various parts of the pine cone crack, whereas the noise and whisper stimuli are more stationary (see Figure 4.5). Due to the short duration of the clicks in pine cones, the loudness percept the participant experienced might be different to what might be expected from the RMS level, due to the temporal integration of loudness (Florentine et al., 1996) and other contributing factors. This was somewhat accounted for by using the time-varying loudness model for assessing the accuracy of judgements; however, this does not permit a rigorous examination of how the temporal envelope affects aural detectability judgements.

The results also showed an effect of stimulus type on the slope of psychometric function, with a steeper slope for noise than for pine cone and whisper. This result might also be explained by spectral-temporal properties of the stimuli. The noise stimulus used in this experiment had a stationary spectro-temporal structure, whereas the pine cone and whisper stimuli had fluctuating temporal envelopes and spectra. The less homogeneous ‘real-world’ stimuli (pine cone and whispers) may have increased the range of levels over which the subject judged the stimulus to be sometimes detectable (p(yes) between 0.1 and 0.9) due to the increased variability in temporal and spectral properties. This effect is analogous to the effect of masker type on the slope of the PF in a speech-based task where more variable maskers (babble, single-talker) increase the range of levels over which the subject can detect some of the signal (MacPherson and Akeroyd, 2014). A similar effect is also known for the slope of the PF, where more homogeneous word lists give steeper PFs (Wilson and Carter, 2001). Another possible explanation for the steeper slope of PF for the noise stimulus may reflect less internal noise for the noise stimuli, where the more homogenous noise stimulus resulted in less random variability in the internal representations of the stimulus compared to the ‘real-world’ stimuli.

The reduced steepness of slope for the ‘real-world’ stimuli may also reflect a more complex decision-making process, as the subject may have more prior knowledge of such stimuli and therefore may use additional parameters to inform their judgements. This more complex decision may manifest itself as a shallower slope of the PF because the additional factors might increase the range of stimulus levels over which the subject judges the sound to be sometimes detectable. This is speculative and requires further investigation, especially as the opposite argument is also plausible; the increased information available (prior knowledge) may aid the decision and result in a steeper slope.

It was hypothesised that judgements would be more accurate for the real-world stimuli than for the noise stimulus (Section 2.3.6.2). This was because prior knowledge is known to affect egocentric auditory distance estimation, so familiarity was expected to aid judgements (Kolarik et al., 2016, Zahorik, 2002). The results do not support this hypothesis the whisper and pine cone stimuli had large errors. We speculatively suggest that, despite the RMS level of whispers being increased to the level of a shout (e.g, approximately 72 dB SPL; Philbeck and Mershon (2002)), subjects were biased by previous experience that whispers are usually only used to communicate over short distances and therefore would not be audible to the target. Data from Brungart and Scott (2001) support this view. They got listeners to judge the perceived distance of talkers, presented via virtual acoustics, using whispered, low-level and high-level speech. Listeners reported distances for the low level and high level speech quite accurately, with an underestimation consistent with other auditory distance perception studies (Zahorik et al., 2005). Listeners reported the whispers to be at a constant distance of 0.8 m at all presentation levels (48 to 82 dB SPL). The authors were unable to explain why these result occurred, but speculated that prior experience of listening to whispered speech under "natural" circumstances restricted their responses to the range of distances that they usually associate with that experience (e.g. someone whispering to them is usually very close). It is possible that the same non-auditory factor affected aural detectability judgements for whispers; subjects inherently associated whispers with nearby communication and therefore determined a target in the distance to not be able to detect it, even at sufficiently high levels. During the experimental breaks, many subjects commented on how they were finding the whisper particularly difficult to judge, and that they believed that it wouldn't travel as far.

A methodological limitation of this experiment in relation to the whisper stimulus type was the amplification of whispers. The same original recordings of whispered speech were scaled to the desired presentation levels, which has poor ecological validity. As such, subjects were being asked to judge sounds that are unlikely to occur in real life. A solution to this would be to use whispered speech that were recorded with a range of vocal efforts, from very low to as high as possible (e.g. 'stage' whispering). This would give a more ecologically valid stimulus.

A further limitation of the method was the use of white noise as an 'unfamiliar' stimulus. It could be argued that many real-world sounds resemble white noise, and so using such sounds does not really permit stimulus familiarity to be explored. Future experiments interested in the effects of stimulus familiarity might prefer to adopt different stimuli, where the spectro-temporal properties are similar, but the stimulus itself it easily identifiable as familiar or unfamiliar. An example would be to compare speech and speech-shaped modulated noise.

5.6.5 Effect of target distance

It was predicted that aural detectability judgements would be sensitive to target distance, such that the subjects would judge sounds to be less detectable as the target distance increased. The locations of the PFs supported this observation. The average increase in PF location between distances was 8 – 19 dB, depending on the stimulus type and target distance change.

Table 5-8 shows the mean RMS level of the location of the PF for each stimulus, at each target distance. The table also displays the mean difference (in dB) between PF locations at each target distance, which indirectly shows the average judged attenuation between each target distance. The theoretically predicted attenuation was computed by calculating the attenuation for each stimulus used in the experiment (n = 10 per stimulus) at each target distance, using the sound propagation model described in Section 4.4.1.1 (ISO 9613-2), with a 60 dB SPL input (at the reference location).

The theoretically predicted attenuation was 6.6 to 8.0 dB. This attenuation is mainly driven by the inverse square law (6 dB per doubling of distance). The additional attenuation is explained by atmospheric absorption and the ground effect; the whispers had the most attenuation due to their mostly high frequency content and the contribution of atmospheric absorption. On average, all conditions had greater judged attenuation than theoretically predicted, suggesting that subjects judged sound to attenuate at a faster rate than theoretically predicted.

Table 5-8 – Mean judged and theoretical attenuation over target distance.

Stimulus		25 m		50 m		100 m
Pine cone	Location (dB SPL)	31.4		40.6		52.0
	Judged attenuation (dB)	*25 - 50 m*	9.2	*50 - 100 m*	11.4	
	Theoretical attenuation (dB)		6.6		6.7	
Noise	Location (dB SPL)	36.0		44.8		56.1
	Judged attenuation (dB)	*25 - 50 m*	8.8	*50 - 100 m*	11.3	
	Theoretical attenuation (dB)		7.2		7.4	
Whisper	Location (dB SPL)	39.0		57.7		70.5
	Judged attenuation (dB)	*25 - 50 m*	18.7	*50 - 100 m*	12.8	
	Theoretical attenuation (dB)		7.7		8.0	

The significant interaction between stimulus type and target distance showed that the effect of target distance was dependent on stimulus type. Post hoc analysis revealed that the interaction was for the whisper stimulus, where the effect of target distance (increase in PF location as target distance increased) was greater for whispers than for noise and pine cones. The stimulus-dependent effect was only present between 25 and 50m, and 25 and 100 m, not 50 and 100 m. This can be seen in

Table 5-8, where the judged attenuation for whispers between 25 and 50 m (18.7 dB) was markedly large than for noise and pine cones, but the difference between 50 and 100 m (12.8 dB) was similar to that for noise and pine cone.

It is possible that the effect (higher judged attenuation for whispers) was present at 50 to 100 m, but the previously discussed method limitation introduced a ceiling effect that might not have allowed the effect to be observed. Recall that in the PF fitting stage of the analysis a high proportion of PFs were incomplete as there was not a stimulus level high enough to yield a p(yes) above 0.5. These cases were given a PF location value of the maximum presentation level plus 5 dB in order to preserve the nature of their responses within the analysis (Section 5.4.3). It is possible that, if higher stimuli levels had been used, the average PF location for whisper at 100 m would have been higher than the value found. A potential explanation for this higher judged attenuation is similar to the previous argument regarding the whisper, in Section 5.6.4. Whispers appear to be influenced by previous experience of whispers being audible at very close distances and this might have biased their judgements (Brungart and Scott, 2001).

A possible explanation for overestimation of the attenuation between each target distance is errors in visual distance perception. If the subject judged the target to be farther away than the true distance, then a higher amount of attenuation would be expected. The egocentric distance estimates do not support this view (see Figure 5.11), where the median estimate at each target distance was underestimated. Because the absolute distances were generally underestimated, one would expect aural detectability judgments to be based on *less* attenuation between each target distance level, not *more.*

The egocentric distance estimates in Experiment 2 were similar to the distance estimates in Experiment 1, though the sample in Experiment 2 underestimated the distance of 100 m target to a greater extent. The method used to gather estimates was slightly different, and the estimates in Experiment 2 were based on only 2 estimates per distance, whereas Experiment 1 had 9 estimates per distance (3 training, 6 experimental trials). Due to the substantially reduced number of estimates, and estimates per distance, it is possible that the within-subject random effects were not reduced in Experiment 2 to the same extent as they were in Experiment 1. There is not sufficient evidence here to raise concerns that the underestimated distance estimates were problematic.

5.6.6 Accuracy of judgements

The accuracy of aural detectability judgements was of key interest in this experiment, as accurate judgements would suggest that someone has good acoustic stealth awareness (ASA). According to the method used here to assess accuracy, and when considering the average PF locations, subjects did not make very accurate judgements. The accuracy got poorer as the target got further away, and the accuracy was particularly poor for whispers (up to 32 dB error).

As discussed in Section 2.3.6.2, decisions that involve computationally challenging tasks, such as the judgements required of subjects in this experiment, are often made through the use of heuristic rules. Heuristics are known to have bias associated with them which can lead to systematic errors (Tversky and Kahneman, 1974). In this experiment, the lack of accuracy, but sensitivity to stimulus level, suggests that a *louder-further* heuristic might have been operating. Further work is required to explore this further.

Accuracy got poorer as distance increased, for all stimulus types. It is possible that this poor accuracy is related to poorly calibrated internal auditory representations of far space (Kolarik et al., 2016). Although not fully understood, it is generally accepted that vision is highly important in calibrating and maintaining the internal representations of auditory space (Lewald, 2013). The mechanisms involved in developing internal auditory space are highly effective for the horizontal localisation of sound sources, as shown in psychoacoustic studies (Blauert, 1997). Relative to the calibration of auditory space for horizontal localisation, little is known about the calibration of auditory space for distance perception, especially in far space. It is possible that the paucity and ambiguity of cues available via audition relative to vision for far space distance perception leads to a reliance upon the visual system for perception of distant objects. This is consistent with the idea that our auditory spatial system is optimised to detect a source and direct the fovea (centre of the gaze) in to line with the source (Perrott et al., 1990).

Supposing that the calibration of our internal representations of auditory space is dependent on exposure to stimuli combined with visual perception of the sources, it is possible, albeit speculative, that the mechanisms involved in developing internal representations of auditory space do not receive sufficient stimuli at far distances as we direct most of our attention to personal or peri-personal space. This is likely to be true for the many stimuli that are rarely encountered in far space, such as pine cones and whispers. Even those stimuli that are experienced relatively frequently in modern society at far distances, such as emergency service vehicle sirens and aircraft, are usually difficult to determine the true egocentric distance of (Wiley et al., 2000).

A significant limitation of analysing the accuracy of judgements using the target sensation level model, and hence a limitation on the conclusions that can be made, is lack of validation of the target sensation level model for the real-world sounds used in this experiment. Any error associated with the model directly affects the validity of the current experimental findings. In order to improve the validity of these findings, a higher degree of confidence in the ‘true’ detectability of the signal is required. The target sensation model is the current method to determine the ‘true’ detectability of a signal. This is made up of two parts, firstly the outdoor sound propagation model (ISO 9613-2) and secondly the time-varying loudness model (Moore et al., 2016). Notwithstanding the previous discussion on the suitability of these models (Section 4.4.2), it was decided that an additional experiment was required to improve the accuracy of the ‘true’ aural detectability thresholds to enable the accuracy of judgements to be investigated more rigorously. Chapter 6 reports this experiment, which aimed to measure the true aural detectability of the sounds used in Experiment 2 behaviourally in order to better assess the accuracy of aural detectability judgements.

5.7 Summary and conclusions

This chapter reported Experiment 2, which measured aural detectability judgements for normal-hearing civilians using the methods developed in Chapter 4. Subjects performed a task which required them to view a distant target in VR in an open outdoor field environment, listen to a sound, and judge whether the target would be able to hear that sound. Psychometric functions (PFs) were measured for each of 9 conditions: 3 sound types (Gaussian noise, pine cones and whispered digits) and 3 subject-target distances (25, 50 and 100 m). Each subject completed these conditions twice over two sessions.

The results showed that judgements were sensitive to the stimulus level, where higher level sounds were judged more detectable than lower level sounds. PFs were fitted and the location and slopes analysed. Judgements, and the trends in judgements, did not differ significantly between sessions. There was an effect of subject-target distance, where sounds were judged less detectable as the target distance increased. Judgements depended on the type of stimulus, with whispers being judged the least detectable, and pine cones the most, at equivalent stimulus levels; put differently, subjects thought whispers were less detectable than noise and pine cones in similar conditions. The effect of target distance was dependent on stimulus type, where subjects judged sounds to become less detectable as target distance increased to a greater extent for whispers than for noise and pine cones. For all stimulus types, people judged sound to decay faster than theoretically predicted. Based on the predictions of the target sensation level model, which predicts the true aural detectability based on sound propagation and loudness models, judgements were generally not accurate. Accuracy decreased as the subject-target distance increased, and was particularly poor for whispers.

The limitations of the methodology and results were discussed. A major limitation was the analysis of the accuracy of judgements. It was decided that to assess accuracy of the judgements without depending on the loudness model, thresholds for each sound would be measured behaviourally, in order to analyse the accuracy of judgements with more confidence.

In summary, the conclusions of this experiment are:

- In normal-hearing civilians, aural detectability judgements are sensitive to the level of the stimulus, with higher levels being judged more detectable to a distant target than lower levels.
- Aural detectability judgements for normal-hearing civilians, measured using the method here, are repeatable..

- As the distance between the subject and the target increased, sounds were judged to be less detectable. This depended on the type of stimulus, though exactly why is not currently known. It may be explained by prior experience of the stimulus and subjects' expectations of that sound, though this requires further investigation.

Chapter 6 Determining the ‘true’ aural detectability of sounds and assessing the accuracy of judgements

6.1 Experiment 3: Introduction

Experiment 2 measured aural detectability judgements for normal-hearing civilians. When considered schematically as shown in Figure 6.1, this refers to the *subject* making a judgement about whether or not the *target* would be able to detect the acoustic event. In experiment 2, the accuracy of aural detectability judgements was assessed using the target sensation level model. This expressed the judgements in terms of the sensation level of the sound experienced by the target, assuming an alert normal-hearing listener. When assessing accuracy this way, judgements in all conditions were, on average, not accurate, with average judgements of the target’s threshold being above the ‘true’ threshold in all conditions. The average error (difference between judgement and ‘true’ detectability) was least for the noise stimuli and greatest for whispers, and increased with increasing target distance.

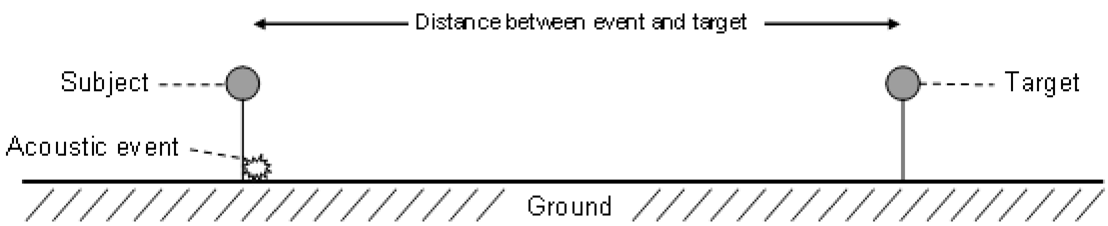

Figure 6.1 - Schematic representation of situations in which aural detectability judgements occur.

A limitation of assessing accuracy by use of the target sensation level model is that the perceptual part of the model, the Cambridge time-varying loudness model (Moore et al., 2016), has not been assessed for the pine cone and whisper stimuli. Whilst the loudness model does take into account temporal integration and frequency-dependent auditory sensitivity, there is the possibility that the model incorrectly predicted the thresholds for the sounds used in Experiment 2. If this is the case, the conclusions of Experiment 2 with regards to accuracy could be incorrect. Due to a primary aim of this PhD being to explore the accuracy of aural detectability judgements, it was considered important that this limitation was addressed in order to increase the confidence in the conclusions.

A solution that would overcome the dependence on the loudness model is to behaviourally measure thresholds for the sounds used in Experiment 2. The ‘true’ detectability of the sounds would then only rely on the sound propagation model, which is largely accounted for by the inverse square law and therefore less debatable. The frequency- and distance-dependent effects of sound propagation (see Section 4.4.1.1 for details) might affect the thresholds for each sound, especially for the whisper and noise, which have a large proportion of their energy at high frequencies. To ensure the behavioural thresholds are appropriate for each distance between source and target, the stimuli used for the behavioural thresholds should include the frequency- and distant-dependent sound propagation effects.

The objective of Experiment 3 was to behaviourally measure thresholds for each stimulus, at each distance, for normal-hearing listeners. These thresholds were then used to provide an estimate of the level at the *subject* location that corresponds to the threshold level of the *target* at each distance, for each stimulus type. This approach removes the use of the loudness model. The estimated levels at the subject location that are equivalent to the target's threshold for each sound then serve as the 'true' aural detectability thresholds. The results from Experiment 2 can then be compared to these 'true' levels to assess the accuracy of judgements.

6.1.1 Summary of aims

The objective of Experiment 3 was to measure the threshold level for the sounds used in Experiment 2. The threshold levels at the target location in Experiment 2 would have depended on the distance that the sound had travelled, due to the high-frequency components of the sound being absorbed to a certain extent during propagation (Wiener and Keast, 1959). Therefore, the thresholds for each sound (pine cone, noise, whisper) at each target distance (25, 50, 100 m) should be measured.

The first aim of Experiment 3 was to estimate the stimulus level required at the subject location to be equivalent to the target's threshold, for each condition (target distance and stimulus type).

The second aim was to combine the results of Experiment 3 (determining the 'true' aural detectability of sounds) and Experiment 2 (aural detectability judgements of sounds) in order to assess the accuracy of aural detectability judgements.

6.2 Methods

6.2.1 Participants

Twenty normal-hearing subjects were used. Normal-hearing was defined as hearing threshold levels less than 20 dB HL up to 8 kHz, as in Experiment 2. The same sample of normal-hearing subjects as for Experiment 2 was used for convenience. It was not considered important whether the sample in Experiment 2 was independent from that for Experiment 3, as there was no reason to expect that prior experience listening to these specific stimuli influences the ability to *detect* the sound. Note, Experiment 3 took place *after* Experiment 2, meaning the participants had not experienced the stimuli before Experiment 2. The sample size of 20 was chosen as this is a comparable sample size to those for other studies estimating thresholds, such as those that contributed towards the ISO 389-1:2017 standard for reference equivalent threshold sound pressure levels (Poulsen and Oakley, 2009, Whittle and Delany, 1966, Poulsen, 2010). These studies yielded 95% confidence intervals of the mean threshold of approximately ± 2 - 5 dB, which was deemed an acceptable error in the estimate for the current study's purpose. It is noted that a sample size of 20 is suitable for experiments where, assuming normally distributed data, the average is the statistic of interest and not the variability between people (Rowan and Pickering, 2011). Participants were paid £30 upon completion of testing.

6.2.2 Apparatus

The experiment took place in the anechoic chamber at the University of Southampton (same testing room as Experiment 2). Stimuli were presented via a single loudspeaker, to recreate the free-field listening conditions that were simulated in Experiment 2. Figure 6.2 shows the simulated situation. The loudspeaker was placed at 0° horizontal azimuth (directly in front of the subject) and adjusted to the height of the subject's head to intersect the direct path of sound from source to receiver. The distance between the loudspeaker and subject was 1.5 m. The subject was seated throughout the experiment for comfort. Although the target in Experiment 2 was standing, the critical factor was that the loudspeaker intersected the direct path of sound from source to receiver. The small difference in loudspeaker height between pine cone and noise/whisper conditions was considered to have a negligible effect, so the loudspeaker was kept at the same height throughout the experiment.

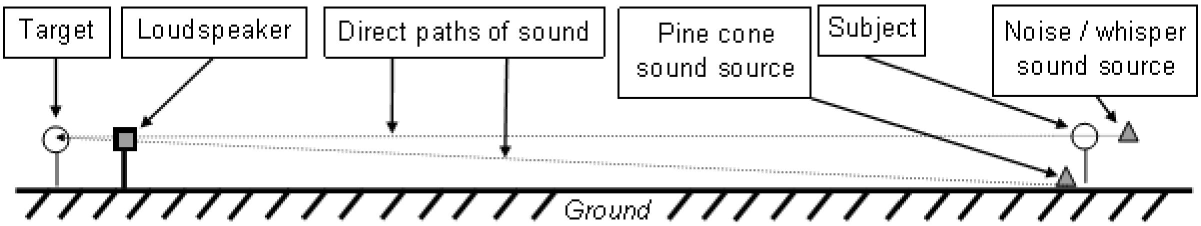

Figure 6.2 - Aural detectability situation. In Experiment 2, sounds were generated at the subject location from the two loudspeakers (triangles). In Experiment 3, the sound sources were reproduced by the loudspeaker (square) and the listener is placed the the target location.

The subject did not wear the VR headset to simulate the outdoor environment. Because the target got no visual feedback about a sound's presence in Experiment 2, it was not important to visually recreate the outdoor environment.

All stimuli were presented using a laptop running MATLAB, and presented via a RME Babyface Pro external soundcard, through a Genelec active loudspeaker. Participants responded via a computer mouse with a graphical user interface button on a computer screen. The laptop and researcher were outside the anechoic chamber during testing. Ambient noise levels were measured prior to the experiment and were within the acceptable limits for soundfield audiometric thresholds (BS, 2009).

6.2.3 Conditions

Recall that in Experiment 2, three stimulus types were used at three target distances. For each stimulus type, ten unique recordings were used to measure the p(yes) at each level in order to measure the psychometric function. Recall also that the ten recordings were selected due to their homogeneity of sound quality and sound pressure levels. Measuring the threshold for each of the ten sound files in this experiment (Experiment 3) would result in a large amount of testing; therefore, it was decided that three recordings from the ten for each sound would be adequate. The three recordings were selected at random for the noise and pine cone stimuli; for the whispers, one recording from each of the numbers was selected at random (1 x "Five", 1 x "Four", 1 x "Nine"). The mean threshold for the three recordings for each stimulus type was used as the 'true' threshold for assessing the accuracy of aural detectability judgements from experiment 2.

Each stimulus type was tested at each target distance (25, 50, 100 m). Target distance was included as a variable because of the distance-dependent effects of the sound propagation model (atmospheric absorption attenuates high frequencies to a greater extent than low frequencies, see Section 4.4.1.1). In total there were 9 conditions (3 distances x 3 stimulus types).

6.2.4 Stimuli generation and calibration

Each raw recording was first processed through the sound propagation model as described in the previous section (4.4.1.1), for each target distance (25, 50, 100 m). Next, all sound files were equalised to have the same RMS value. The RMS value that all stimuli were equalised to was selected to give the maximum signal magnitude whilst avoiding clipping. This had the effect of spectrally shaping the files as if they were generated and received at the source and receiver locations.

For calibration purposes, a 1/3 octave band of noise with a centre frequency of 1 kHz was generated to have the same RMS value as the stimuli (calibration noise). A calibrated microphone connected to a sound level meter was then placed at the reference point, which was the location of the centre of the subject's head without the subject present (the chair was present). The loudspeaker then presented the calibration noise and the loudspeaker gain control was adjusted until the level of the calibration noise was 40 dB SPL at the reference point. This level was equivalent to a maximum level during the experiment. Stimulus level was then controlled in MATLAB by scaling the stimulus using the following formula:

$$Signal_{Scaled} = Signal_{40} \times 10^{\frac{D-M}{20}}$$

Where: $Signal_{40}$ is the equalised 40 dB SPL sound file, $Signal_{Scaled}$ is the sound file scaled to the desired level, D is the desired stimulus level (in dB SPL) and M is the maximum stimulus level, which in this experiment was 40 dB SPL.

6.2.5 Threshold measurement

Thresholds were measured using an automated adaptive procedure, delivered via MATLAB routines. The MATLAB code was written by Daniel Rowan. The procedure was an automated version of that used to measure pure tone thresholds in clinical audiometry (ISO 8253-1:2010). The subject was instructed to sit quietly and actively listen for a sound, and press the response button as soon as they heard a sound. The starting level was randomised between 20 and 40 dB SPL (to the nearest 1 dB). The sound was presented to the subject and then the computer waited for 1 second for a response. If the subject responded during the stimulus or in the 1-second period after the stimulus, the trial was deemed a hit. If no response was recorded, the trial was deemed a miss. The subject was not given any warning of when a trial was about to start or any feedback about whether they correctly identified a sound. If the trial was a hit, the stimulus level of the next trial decreased by 10 dB. If the trial was a miss, the stimulus level of the next trial increased by 5 dB. The inter-trial duration was randomised between 1.5 and 3.5 s. False positives (responses outside of the response period) were disregarded. The adaptive track continued until two hits on ascending runs at the same level occurred. The level at which the two ascending hits occurred was taken as the threshold for that sound file. This was conducted for all 27 sound files (9 sound files at 3 distances).

6.2.6 Procedure

Participants first underwent otoscopic examination to check that their ears were disease-free and not occluded by wax. Next, their pure tone hearing threshold levels were tested using a calibrated audiometer to check that their thresholds met the inclusion criteria. Finally, a hearing questionnaire checked they had not had any recent noise exposure, tinnitus or recent ear-related problems.

Next, participants were instructed and the loudspeaker was adjusted to the height of the participant's head whilst they were seated. A practice threshold measurement then commenced to give the participant an opportunity to learn the procedure. These results were not analysed. Threshold testing then commenced. The 27 thresholds were organised into 3 blocks; each block was for one target distance. The order of the three blocks was counterbalanced, such that each possible order (6 orders) had equal numbers of participants, apart from participants 19 and 20 who had random orders. In each block, each threshold for the 9 sound files was carried out in a randomised order. A MATLAB function was used to randomise the order. Participants had a 5 to 10-minute break after each block of 9 thresholds.

6.2.7 Calculation of 'true' detectability

As aforementioned, the mean stimulus level of the three thresholds in each condition was taken as the target threshold level for that condition. This threshold level represents the RMS level at the reference point (location of target's head in the absence of the target) that was equivalent to the target threshold in the respective condition. In terms of the schematic diagram in Figure 6.3, the target threshold level is the level of the acoustic event at the target location that the target can just detect.

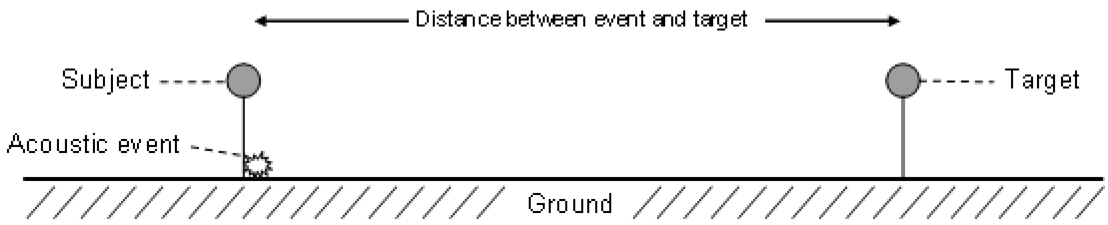

Figure 6.3 - Schematic diagram of aural detectability judgement.

The first aim of Experiment 3 was to estimate the level at the subject equivalent to the target threshold level. Following measurement of the target threshold level for each condition, the stimulus level at the subject location that corresponded to the target threshold level was calculated for each stimulus type and target distance. This calculated level serves as the ‘true’ detectability level, to which the aural detectability judgements can be compared in order to assess the accuracy.

To achieve this, an adaptive process, incorporating the distance between source and target, the sound propagation model and the target threshold level was utilised, as shown in Figure 6.4. The outdoor sound propagation model implement as described in Section 4.4.1.1 and shown in Appendix B. The procedure adapted the level at the subject (input level) until the level at the target was within 0.1 dB of the target threshold.

This experiment was approved by the University of Southampton Ethics and Research Governance board (Ethics number = 23783.A1).

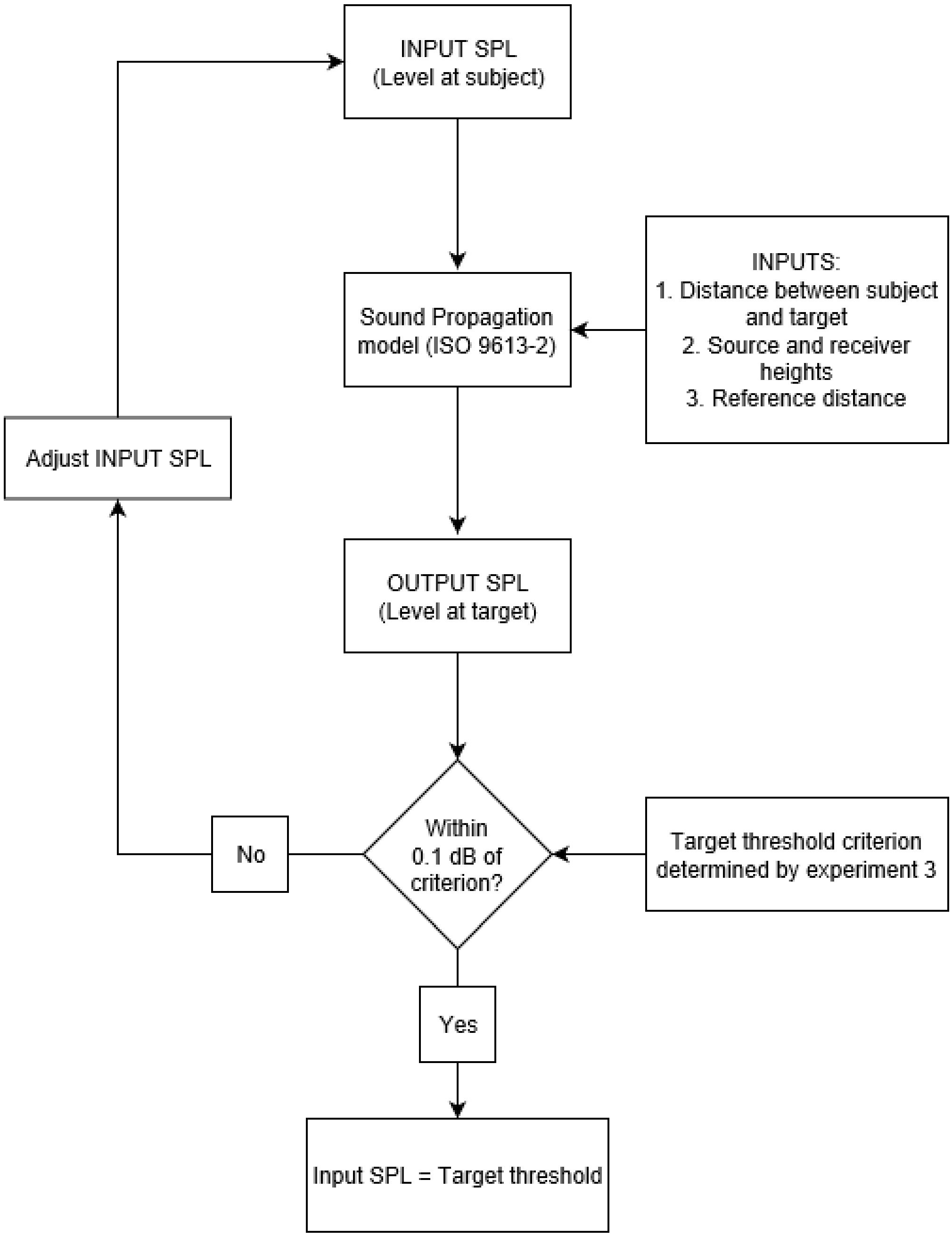

Figure 6.4 - Adaptive process for calculating the level at the subject equivalent to the target threshold level.

6.3 Results part 1: 'True' aural detectability levels

The results for each condition were first checked for normality. No condition yielded a significant Shapiro-Wilk test statistic and box plots for all conditions appeared normally distributed. The mean thresholds at the target and standard deviations for each stimulus at each distance are shown in Table 6-1. The corresponding level at the subject is also shown in the table. Thresholds are reported in terms of their RMS level.

Table 6-1 - Experiment 3 results for each stimulus type and target distance. The 'true' detectability level is shown for each condition, displayed as the level at the subject and the target that corresponds to the average target threshold. The standard deviations (SD) of the behaviourally measured target thresholds are shown for each condition.

Condition		**'True' detectability level**		
Stimulus type	***Target distance (m)***	***Level at subject (dB SPL)***	***Level at target (dB SPL)***	***SD (dB)***
Pine cone	25	24.5	-0.3	3.9
	50	30.7	-1.3	3.7
	100	38.3	-1.1	3.4
Noise	25	32.7	4.4	3.9
	50	40.1	3.2	3.6
	100	48.4	2.5	3.8
Whisper	25	33.2	2.1	3.9
	50	38.6	1.8	3.9
	100	44.7	2.1	3.8

6.4 Results part 2: Accuracy of aural detectability judgements

Following the completion of Experiment 2 (aural detectability judgements) and Experiment 3 (determination of 'true' detectability), the results were combined to assess the accuracy of aural detectability judgements. For this, the mean level at the subject location that corresponded to the location of the PF (from Experiment 2) was used to represent the aural detectability threshold judgement, for each stimulus type (pine cone, noise, whisper) and each target distance (25, 50 100 m). This level was compared to the 'true' detectability level, although it is acknowledged that this may not be true due to the dependence on the sound propagation model. The target sensation level model was also compared, in order to assess how the model compares to the 'true' detectability. These are shown in Figure 6.5 for each stimulus type and each target distance. Note, the error bars for the 'true' detectability thresholds show the 95% confidence interval and were calculated using standard error of the mean of the behavioural threshold, and reflect the variability in the sample of twenty normal-hearing listeners.

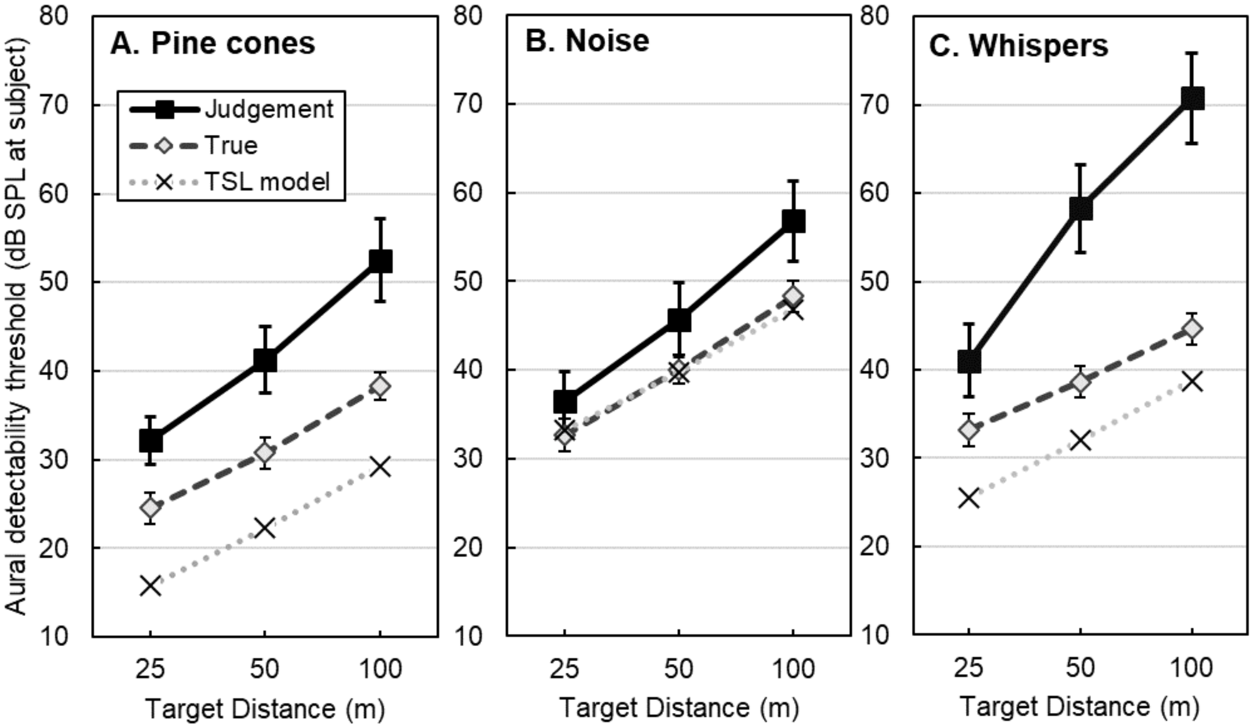

Figure 6.5 - Aural detectability thresholds for each stimulus type and target distance. Panel A is for the pine cones, B for the noise and C for the whispers. The judgements from Experiment 2 are shown by the black squares (Judgement). The 'true' thresholds, according to Experiment 3 (behavioural thresholds with sound propagation model), are shown by the grey diamonds (True). The crosses show the aural detectability thresholds according to the target sensation model (TSL). Each threshold shows the stimulus level at the subject. The error bars for the judgements and true thresholds show the 95% confidence interval of the mean.

In all conditions, the mean aural detectability judgements were not accurate. The mean level that subjects judged to be at the target's threshold was above the 'true' threshold. This means that, on average, when the subject judged the sound to be inaudible to the target, it was audible.

The error (difference between 'true' and judgement) was calculated for each subject for each stimulus type and target distance and shown in Figure 6.6. A repeated-measures ANOVA with stimulus and distance as factors showed significant main effects of Stimulus, $F(1.428, 38.555) = 35.079$, $p < 0.001$ (Greenhouse-Geisser correction), and of Distance, $F(1.461, 39.457) = 73.716$, $p < 0.001$ (Greenhouse-Geisser correction), and a significant interaction between Stimulus and Distance, $F(2.213, 59.761) = 40.834$, $p < 0.001$ (Greenhouse-Geisser correction). Pairwise comparisons with Bonferroni corrections showed that the mean error increased as distance increased for all stimuli at all levels, except for the noise stimuli between 25 m and 50 m. The error smallest for the noise stimulus, then the pine cones, then the whisper. The whisper stimulus resulted in the largest error (on average), with a mean error of approximately 26 dB at a target distance of 100 m.

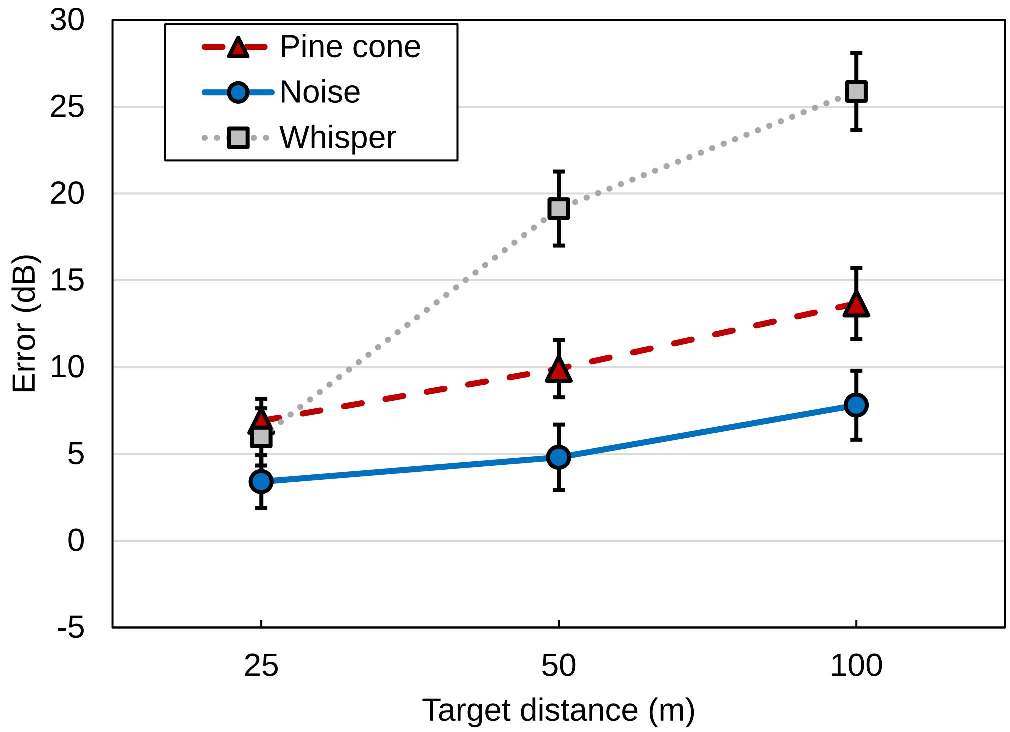

Figure 6.6 - Mean error of aural detectability judgements in Experiment 2. Error bars show the 95% confidence interval of the mean.

6.5 Discussion

6.5.1 Overview

Experiment 3 behaviourally measured thresholds for the sounds in Experiment 2, to enable an analysis of the accuracy of judgements that did not depend on a loudness model. This discussion starts with a comparison of the behavioural threshold to those predicted by the loudness model (Section 6.5.2). Then, the accuracy of judgements is considered (Section 6.5.3), followed by an exploration of the distribution of errors in judgements for each condition (Section 6.5.4). This chapter ends with a summary of conclusions.

6.5.2 Comparison of behavioural to loudness-model-predicted thresholds

The predictions of target threshold levels for each condition made by the TSL model were compared to the behaviourally measured thresholds in order to evaluate the TSL model's accuracy. Since the 'true' detectability levels measured in Experiment 3 and the detectability levels predicted by the TSL model differ only in terms of the perceptual aspect, any difference between the two approaches can be attributed to differences between the time-varying loudness model and behavioural thresholds. The differences between the 'true' thresholds and the TSL model thresholds can be seen in Figure 6.5, as the difference between 'True' and 'TSL model'.

The TSL model very accurately predicted the 'true' aural detectability levels for the noise conditions, with differences between mean behavioural and loudness-model-predicted thresholds of less than 1.5 dB. This was expected due to the variety of comparison data that the model was shown to give accurate threshold predictions for (Moore et al., 1997), as discussed in Section 4.4. The close agreement between the thresholds measured here and the loudness model suggests that the method and calibration used were effective.

The TSL model was less accurate for the 'real-world' sounds. The behaviourally measured thresholds were, on average, 8-9 dB and 6-7 dB higher than the loudness-model-predicted thresholds for pine cones and whispers, respectively. The errors were consistent at each target distance. These errors are similar to those reported by Rennies et al. (2013), who found errors between behaviourally measured loudness matches (and categorical loudness scaling) and predictions by the same model used here, of up to 15 dB. However, these data were for loudness matching complex stimuli (speech-like and 'technical' sounds) to a 70 dB SPL reference signal, so these data do not allow a direct comparison to the threshold data measured here. Analysing the specific components of the model in order to account for the measured error is beyond the scope of the current work. Although the model is designed to mimic the temporal integration properties of the human auditory system, it is possible that the time constants used in the model were inappropriate for the stimuli used here.

Due to the errors observed for the ‘real-world’ stimuli, the TSL model is not recommended in its current form unless previously validated sounds are of interest; should such a model be required, future work should either replace the perceptual component of the model (loudness model) for behaviourally measured thresholds, or invest time in improving the loudness estimate for the specific stimuli of interest.

6.5.3 Accuracy and variability of judgements

The results of Experiment 3 were combined with those of Experiment 2 to allow the accuracy of judgements to be assessed. Figure 6.5 shows the aural detectability threshold, which is the level at the subject location that corresponds to the threshold level at the target location. The figure shows:

1. Judgement: the judged aural detectability threshold. This uses the average PF location of the judgements made in Experiment 2.
2. True: the ‘true’ aural detectability threshold, based on Experiment 3.
3. TSL model: the TSL model prediction of aural detectability threshold.

As discussed in the previous section, the TSL model agreed with the ‘true’ aural detectability thresholds for the noise conditions, but was up to 9 dB off for the pine cone and whisper conditions. As a result, the findings relating to the accuracy of judgements in the previous chapter need revising.

Despite the errors in the TSL model, the trends in the accuracy of judgements remain unchanged: accuracy decreased as the target distance increased; judgements were most accurate for noise stimuli, followed by pine cones, then whisper; judgements became markedly less accurate with increasing distance for the whispers. The magnitude of error in judgements was reduced for the pine cone and whispers due to the error in the TSL model, but large errors in judgements were still observed.

The familiar stimuli (pine cones and whispers) were expected to be judged more accurately due to the prior experience of those sounds. The data showed that these stimuli were judged less accurately than the non-familiar sound (noise). However, given the confound of spectro-temporal differences between stimuli, it is not possible to conclude whether or not stimulus familiarity can explain the differences in accuracy.

The variability in the accuracy of judgements requires consideration. Although the experiment was designed to measure the effects of basic factors on aural detectability on average, it is of interest to explore the variability of the sample. The standard deviations (SD) of the thresholds for each stimulus recorded in Experiment 3 (behavioural thresholds), 3.5 – 4 dB, were substantially smaller than the SDs of the locations of the PFs recorded in Experiment 2 (aural detectability judgements), 6 – 13.5 dB. It is likely that the difference in variability reflects the nature of the task.

In Experiment 3, the listener had to listen and respond when they detected the presence of a sound; no further cognitive processing was required. The sources of variability for this task include between-subject differences in hearing ability, within-subject differences in decision criterion for when to respond to a sound, and random factors. In contrast, the task in Experiment 2 involved the subject listening to the sound and making a substantially more complex judgement based on their perception of that sound. Therefore, two additional sources of variability were present, the individual's aural detectability decision criteria and all the 'higher level' processing involved, and the visual distance perception. The sample for Experiment 3 (n = 20) was drawn from the sample in Experiment 3 (n = 28), so differences in hearing ability can be excluded in this comparison. The findings of Experiment 1 suggested that, on average, distance estimation was similar in VR and reality, but that there was large variability between subjects. The large between-subject variation in aural detectability judgements in Experiment 2 could potentially be explained by variability in distance estimate. However, the amount of variability in distance estimation (SD = ~30% estimation error) cannot fully explain the large variability observed in judgements (SD = ~10 dB). Therefore, it is likely that variability in aural detectability decision criteria also contributed to the observed variability. The reason aural detectability decision criteria vary amongst people is not clear.

6.5.4 Distributions of judgement error

Given that there was large variability in this sample in terms of the accuracy of their judgements, it is appropriate to assess the accuracy on an individual level by viewing the distribution of data rather than just the average. The difference between the judged aural detectability threshold (location of PF) and the 'true' detectability was calculated. This value (in dB) represents the error in judgement, where a negative value shows that the judgement was below the target's threshold (inaudible to the target) and a positive value shows that the judgement was above the target's threshold (audible to the target). The distributions of error for each condition are shown as histograms in Figure 6.7.

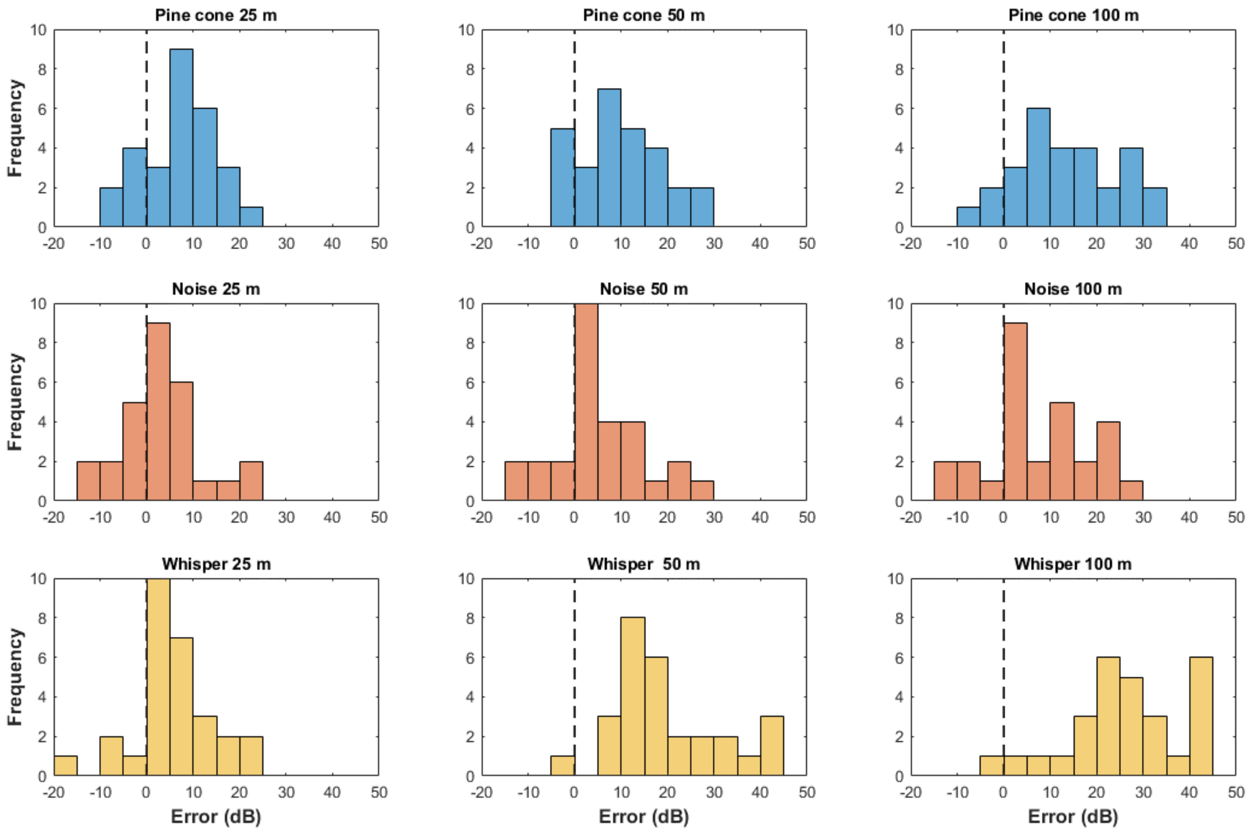

Figure 6.7 - Histograms of errors in aural detectability judgements for each condition. Results for pine cones, noise and whispers are shown in blue, orange and yellow, respectively. The target distances are 25, 50 and 100 m from left to right. The vertical dashed black line at 0 dB represent the ‘true’ detectability (target threshold).

The histograms show that in all conditions, the majority of subjects had positive errors, which means their judged threshold was above the target’s threshold. In each condition, a small number of participants gave judgements that had negative error. For ASA, it is important that aural detectability judgements have 0 dB, or less, error, as this means the individual is aware that their stealth may be compromised.

To explore whether a subject was in a similar part of the distribution in each condition, a Spearman’s rank correlation matrix was calculated. If a significant and strong correlation between a pair of conditions is found ($\rho > 0.7$), this suggests the patterns of errors in those two conditions is very similar. Table 6-2 shows the correlation matrix.

Table 6-2 - Spearman's rank correlation matrix for ranked error. All correlations were significant at the p < 0.01 level, excluding those in grey shaded boxes where $p > 0.05$, and those with an asterisk where $0.01 < p < 0.05$.

	P25	P50	P100	N25	N50	N100	W25	W50	W100
P25									
P50	0.91								
P100	0.80	0.92							
N25	0.62	0.68	0.51						
N50	0.68	0.85	0.74	0.80					
N100	0.68	0.86	0.84	0.71	0.91				
W25	*0.35*	0.49	0.46*	0.68	0.68	0.63			
W50	*0.36*	0.57	0.58	0.62	0.73	0.76	0.86		
W100	*0.25*	0.45*	0.46*	0.58	0.62	0.61	0.86	0.93	

The correlations show that for the majority of conditions, except between pine cone at 25 m and the whisper conditions, the ranked error was moderately or strongly related. This suggests that subjects tend to be in similar parts of the distribution in each condition, for example, the least accurate subjects overall tend to be least accurate in each condition.

The correlations are weakest between scores for pine cones and whispers (correlations at the bottom left of the matrix), suggesting that the errors people made were not consistent between the pine cones and whispers. It could be that the non-auditory factors that influence judgements (e.g. stimulus familiarity), were different for different people. This is highly speculative and requires further investigation.

6.5.5 Generalisation

The findings in terms of the errors in aural detectability judgements must be considered in terms of their generalisability: to what extent would an individual make similar judgements outside of the laboratory? For the specific task that was simulated here, we argue that the identified trends are likely to be representative of real-world performance. The target distance was important for judgements, so it is important that distance perception cues were available in the VR environment. Distance estimation was found to be similar, on average, in VR as in reality over the ranges that we tested (Experiment 1), so the trends in the results here should not be affected by dissimilar distance perception.

The acoustic environment (anechoic conditions) was generally perceived by most subjects as strange, but under the appropriate meteorological conditions and ambient background noise levels, a real outdoor environment would approximate quite closely the acoustic environment experienced by the subject.

Performing the task successively many hundreds of times may have affected the judgements, such that the results do not reflect what might occur for a single aural detectability judgement, made outside of an experimental context. The results from session 1 and session 2 in Experiment 2 were similar, suggesting that the judgements people made, on average, were repeatable during independent sessions. This suggests that if there was a training or learning effect, it was not measurable across the two sessions.

A limitation to the result for the whisper stimuli is unlikely presentation levels. If subjects were biased by their previous experiences of whispers, such that they expected a whisper to have constant source power, a loudness constancy effect may have occurred (Zahorik and Wightman, 2001). If this was the case, the increased level for whispers would have been experienced as the source getting closer to them rather than the source power increasing. Data from Brungart and Scott (2001) support this view, as they found that subjects reported the egocentric distance for whispers to decrease as the presentation level increased, despite the simulated source distance remaining constant. Such an interpretation could explain why some subjects never responded “Yes, the target would be able to detect that sound” even at the loudest presentation level. The inconsistency between subjects suggests that if loudness constancy was present, it was not present for all subjects (Figure 6.7). Further research is required to explore this effect in more detail. Should this work be carried out, the stimuli should take into account varying vocal effort whilst controlling the stimulus level, to ensure the whispers are more ecologically valid than those used here (see Section 5.6.4).

6.6 Summary and conclusions

This chapter reported Experiment 3, which behaviourally measured the thresholds for the sounds used in Experiment 2. By combining the results of these experiments, the accuracy of judgements was assessed. The data in these experiments suggest that most, but not all, people make errors in their judgements of whether or not a target could aurally detect them. The direction of the error was such that they judged the sounds to be undetectable to the target when, according to the methods used here, they would be detectable. The magnitude of the error was variable, and depended on the target distance and the stimulus type. Precisely why variability between people exists is not clear, but it seems that non-auditory factors, such as prior experience of the sounds, might have affected the judgements. Given the complexity of the task (inferring the physical characteristics of a sound at a different location and inferring its detectability), it is likely that heuristic rules guided people's judgements. Such heuristic rules are subject to a range of biases (Tversky and Kahneman, 1974), which might explain why errors occurred.

Based on the results of Experiment 3, the following conclusions were made:

- The Cambridge time-varying loudness model (Moore et al., 2016) predicted the absolute threshold of Gaussian noise accurately, to within 1.5 dB of the behaviourally measured threshold; the model was less accurate for the real-world sounds, with errors up to 9 dB. The use of the loudness model is recommended for previously validated sounds, but not un-validated 'real-world' sounds, when modelling the aural detectability of sounds.
- Aural detectability judgements in normal-hearing civilians, for the subject-target distances and environment simulated here, are generally not accurate, with most subjects reporting an audible sound as inaudible to a distant target.
- The error in judgement is greater as the distance between the person and the distant target increases, over 25 – 100 m subject-target distances.
- The accuracy of judgements was variable between people, with some people giving highly accurate judgements, , but most giving inaccurate judgements. The components that contribute to this between-subject variability are unknown, but potentially involve non-auditory factors such as the individual's aural detectability decision criterion.
- Non-auditory factors affect the judgements people make, especially for sounds that are usually heard at quiet levels and at short range. The exact nature of this is currently unknown and requires further investigation.

Chapter 7 General discussion

The three experiments reported in this thesis have contributed towards a greater understanding of peoples' ability to judge whether or not sound sources close to them are detectable to a distant target. A method was developed that allowed aural detectability judgments, hypothesised to be how ASA can be achieved, to be measured and was used to measure aural detectability judgments for normal-hearing civilians. Experiments 2 and 3 showed that aural detectability judgements, for normal-hearing civilian participants, are repeatable, sensitive to target distance, sensitive to stimulus level, and depend on the type of stimulus. It was also shown that the accuracy of judgements varies across participants, with some participants giving accurate judgments, but most giving inaccurate judgements. Accuracy was worst for whispers, and it was hypothesised that prior experience of whispers (that they are usually quiet and perceived at short distances) biased judgements. How this work fits into the broader picture of ASA and AFFD is discussed in this chapter, including the implications for the military, suggested future work required for the military and how hearing acuity might affect the task, and how it could be explored.

7.1 Methodological improvements

The novel method developed in this PhD project was shown to have, to some extent, construct validity, as people's judgements were sensitive to stimulus level and target distance. Although this observation gives some information about construct validity, ideally a study that carries out the same task in a real-life environment, similar to the virtual environment used here, would help build confidence in the laboratory-based method's ability to measure aural detectability judgements with ecological and construct validity.

Whilst the results of this study provide an interesting insight into aural detectability judgements, it is recommended that methodological improvements are made prior to further experimental work. The most obvious change required to the novel method used in Experiment 2 is the method for selecting the stimulus levels. Greater variability between subjects was observed than expected based on the pilot study, and the familiarity period that was used at the start of the testing session did not identify the appropriate set of stimulus levels with total success. A potential solution to this problem might involve having a familiarisation period so that the subject can experience the task first, followed by a method of adjustment based task (as used in the very first pilot, Section 2.4.2). The method of adjustment task would involve the subject adjusting the level of the stimulus to the level they think is just detectable by the target. The set of stimulus levels could then be spaced out around that identified level, thus maximising the chance of measuring the full PF.

Given the significant recruitment challenge involved in testing military, and military hearing impaired participants, another potential improvement would be to reduce the number of trials required per condition. A great amount of testing was required due to the use of the method of constant stimuli. Unless the aim of future studies to explore the full psychometric function in detail, adaptive procedures are recommended. This would reduce the time per condition, enabling more factors to be tested or more repeats to be measured. Measuring aural detectability judgments with greater efficiency will allow more factors to be explored. Factors that could be explored using the VR approach are include:

- Background noise, using ambisonics or virtual acoustics
- Vocal effort (shout, conversational level, whisper)
- Technical sounds (Velcro, weapons manipulations, radio noise)
- Spectral and temporal characteristics

7.2 Generalisation to the military

Based on the results of the experiments conducted in this study, in very quiet situations, people without explicit acoustic stealth training can make large errors in aural detectability judgements. This has potential implications for close-combat personnel operating in acoustic stealth situations. Further research is required to understand if such erroneous judgements occur for military personnel.

A significant limitation of the generalisability of the results in this thesis with regard to someone's ASA is the ecological validity of the situation. The experimental design employed in this study was based on the hypothesis that an aural detectability judgement takes place in order for someone to assess their aural detectability with respect to a nearby enemy. As was shown, a large amount of variability between subjects was associated with judgements, and non-auditory factors appeared to affect judgements. It is therefore possible that the results of this experiment do not generalise well to a soldier in a real life-or-death acoustic stealth operation. If the consequences of poor performance are severely negative (e.g. death), the subject might be more inclined to report a sound as detectable, compared to the situation such as that simulated in the experiments here. It is predicted that the likely outcome of judgements framed more closely towards real-life acoustic stealth situations is an increase in the false alarm rate. Subjects will probably be more cautious before saying the target could not detect that.

An additional factor not considered in this experiment is the role of multi-tasking. A soldier will never be exclusively judging their aural detectability whilst conducting a military operation. Therefore, it is possible that ASA is difficult to achieve, because of attentional resources being allocated to other tasks (e.g. communication, map-reading, visual task). The methodology employed does not simulate a multi-tasking situation well, though the almost limitless possibilities of VR mean it would be feasible.

The environment that was simulated in the current experiments also represents an unrealistic situation for soldiers to be operating in. Soldiers will always adopt a path that permits visual stealth as well, meaning that the acoustics will be complicated by acoustic barriers and reverberation. The environment used here was selected for its relative simplicity for modelling the acoustics and on the basis that visual distance estimation was similar in VR to reality in this environment. Therefore, whilst the results here show interesting trends, it is possible that soldiers would never find themselves making judgements in such acoustic environments (free-field).

7.3 Implications for auditory fitness for duty

This study was primarily motivated by the presumed role of hearing in ASA. Although not clear how the results here will generalise to acoustic stealth behaviour, it is possible that without experience or training soldiers could make errors whilst operating in situations requiring acoustic stealth. The aural detectability judgements were sensitive to stimulus level, which suggests that hearing has an important role. Hearing loss could therefore affect a person’s aural detectability judgements due to reduced auditory feedback of the acoustic event.

As seen in studies investigating the effect of hearing loss on operational performance (Keller et al., 2017), a hearing-impaired soldier might employ compensatory strategies in order to overcome the challenges imposed by reduced auditory feedback. Chapter 2 suggested that aural detectability judgements could occur *a priori* or *post hoc*. It is possible that *a priori* aural detectability judgements, in combination with other sensory modalities, allow hearing-impaired soldiers to operate effectively in acoustic stealth situations. For example, by having prior knowledge of acoustically intense actions (e.g. movements, loading rifle), they are able to avoid making aurally detectable levels of sound. They might also use feedback from other sensory modalities, such as visual and tactile information, to make aural detectability judgements and overcome the lack of auditory feedback imposed by hearing impairment.

7.4 Hearing protection devices

It is a mandatory requirement that soldiers wear hearing protection devices (HPDs), though it is known that soldiers often do not (Bevis et al., 2014, Okpala, 2007). Common reasons for not wearing hearing protection are the lack of situation awareness due to reduced ability to detect auditory warnings (Abel, 2008). It is highly probable that HPDs will interfere with ASA, as shown by Casali et al. (2009). The effect of HPDs on ASA should be considered, given the mandatory requirements designed to prevent noise induced hearing loss. Although beyond the scope of this project, hearing protection devices are a critical component of maintaining a soldier's AFFD by reducing their noise exposure (Nakashima and McDavid, 2018). If HPDs impact ASA, this may cause a soldier to remove their hearing protection. Hence efforts should be made to identify the problem and engineer solutions to overcome this.

7.5 Recommendation for field studies

Due to aforementioned limitations in laboratory aural detectability judgements (e.g. framing effect, unlikely situation), a field study that measures the operational impact of hearing loss on acoustic stealth behaviour would complement laboratory studies. In order to investigate the effects of hearing loss and hearing protection devices, a field study design, similar to those implemented by studies discussed in Chapter 2, would be appropriate (Sheffield et al., 2017, Sheffield et al., 2016a, Sheffield et al., 2016b, Casali et al., 2009). This would involve a task for a group of soldiers that requires acoustic stealth. To explore the impact of hearing loss, real-time hearing loss simulators could be used to manipulate the available auditory feedback in a manner similar to hearing loss. As discussed in Experiment S3, masking noise is required to mask the direct-path of sound for real-time hearing loss simulators (Section A.4). To explore the effect of HPDs, the user could complete the task with and without HPD. Objective measures of acoustic stealth could be performed by use of calibrated microphones in various locations, such as each soldier's helmet and at the enemy locations. Analyses of acoustic output (from the helmet microphone) and aural detectability (from the enemy location microphone) could then be carried out to measure the acoustic impact of hearing loss. Additional operational measures based on the task, such as time to complete, GPS tracking and self-report measures could be used to measure the operational impact, such as those used by Sheffield et al. (2017).

An additional measure, which has so far only been used in one AFFD-related study, is eye-tracking data (Keller et al., 2017). As discussed in Section 2.2.3, soldiers might employ compensatory strategies to overcome the challenges imposed by hearing loss, such as greater reliance on visual cues (Keller et al., 2017). Portable eye-trackers are available for tracking the gaze of a user (Mele and Federici, 2012). Using these techniques, one could analyse if a soldier modifies their visual search or gaze behaviour in response to manipulations in auditory feedback (e.g. hearing loss simulation, HPD).

Such an experimental design would be highly valuable in assessing the impact of hearing loss on operational effectiveness in stealth situations, but it would be challenging to achieve. Firstly, the practical, financial and logistical arrangements would be very challenging for civilian researchers, so such a study is only recommended for researchers with access to appropriate facilities and personnel. Secondly, the data would be subject to the conditions during data collection and therefore out of control of the experimenter.

Should field studies reveal that ASA is poor despite military training, explicit training will be required. Training in the context of aural detectability judgements could include: 1) simple rules, and 2) feedback on task performance. Soldiers could be given general rules (heuristics) about the range of aural detectability of common sounds, such as quiet walking, whispering, stepping on twigs and cocking rifles. This would reduce the effect of the potentially erroneous aural detectability decision criteria that, according to the experiments here, could lead to errors in judgements that have fatal consequences. Training could also include feedback from experts that allows soldiers to learn if they are behaving too loudly for the specific situation they are in, thereby reinforcing successful stealth behaviour.

Chapter 8 Conclusions

8.1 Conclusions

This study investigated aural detectability judgements by normal-hearing civilians. This was motivated by the potential importance of hearing during situations requiring acoustic stealth awareness (ASA), and how hearing loss could negatively impact an individual's auditory fitness for duty (AFFD). Few studies on ASA were identified. It was suggested that ASA is at least partly achieved by the subject (the soldier) judging whether a sound they make, or a sound made nearby, is audible to a distant target (the enemy). This is referred to as an aural detectability judgement. It was unclear whether people can consistently make aural detectability judgements, how accurate they are and the extent to which the judgements are affected by various factors, such as the subject-target distance and the intensity of the sound. Due to the substantial gap in knowledge, a novel method was developed that required subjects to view a distant target in virtual reality (VR), listen to a sound and judge whether the target would be able to aurally detect that sound.

The conclusions of this PhD project are as follows:

- Visual distance estimates made in a VR environment were usually within approximately 10% of estimates made in the corresponding real environment for distances between 25 and 125 m, when using a large open outdoor field environment during the daytime containing multiple distances cues, a human-sized target and an Oculus Rift (CV2). There was a trend for distance to be underestimated in VR relative to reality for distances of 75 – 125 m, though this was not statistically significant. It remains unclear as to whether VR is suitable for applications requiring distance estimation of human-sized targets beyond 125 m in similar environments to that used here.
- Using the developed method, normal-hearing civilians made aural detectability judgements that were fairly stable between two sessions completed on separate days. Aural detectability judgements were also sensitive to the subject-target distance, and the level of the sound, which in combination with the repeatability of judgements suggests the method has some construct validity. This method provides an approach to studying ASA that has a high degree of practicality and experimental control, and should be used for future studies of ASA. Future research should compare this laboratory method to some similar conditions in real outdoor environments in order to further establish the construct validity and increase the confidence in using the method to further explore ASA.

- Aural detectability judgements were sensitive to sound level and subject-target distance, at least for normal-hearing civilian subjects. Higher sound levels were judged to be more detectable than lower sound levels, and the detectability of a sound decreased as the subject-target distance increased. These effects are consistent with theoretical predictions based mostly on the inverse square law.
- Aural detectability judgements for Gaussian noise, the crunching of pine cones and whispered digits changed with increasing subject-target distance at a rate faster than expected from theoretical predictions, at least for normal-hearing subjects. Put differently, people over-estimated how rapidly the amplitude of the sounds would decay with distance. This was particularly strong for whispered digits, which might indicate that people’s prior expectations about the level of a sound (e.g. whispers are expected to be quiet) influence aural detectability judgements along with the level of actual level of the sound. Further research is required to confirm that prior expectations play a role in aural detectability judgements.
- The Cambridge time-varying loudness model (Moore et al., 2016) accurately predicted the average absolute thresholds of normal-hearing listeners for Gaussian noise to within 1.5 dB. The prediction of absolute thresholds for the crunching of pine cone and whispered digits was less accurate, but still within 9 dB of behaviourally measured thresholds. For modelling the aural detectability of the sound in ASA, the use of the loudness model is appropriate for previously validated sounds, but less so for real-world noises. Further research is required for the prediction of absolute thresholds of real-world sounds should this model be used for modelling the aural detectability of non-validated sounds. Behaviourally measuring thresholds for these sounds is likely to be more accurate.
- On average, aural detectability judgements by normal-hearing civilians are inaccurate, with most subjects reporting a sound predicted to be audible to a distant target as being inaudible, at least under the conditions and predictions used here. This is not due to the potential inaccuracy in visual distance perception (e.g. due to VR) noted above since that would have made targets appear closer than they were, which presumably would have tended to make subjects judge sounds to be more, not less, detectable. On average, the magnitude of error in aural detectability judgements increased as the subject-target distance increased. The error was least for Gaussian noise, and markedly worse for whispers, especially at a subject-target distance of 100 m.
- Aural detectability judgements in normal-hearing civilians varied quite substantially between people within a given condition (subject-target distance and sound type). The least accurate subjects overall tend to be amongst the least accurate subjects

in each condition, and the most accurate subjects overall tend to be the most accurate in each condition.

8.2 Future directions

The recommended future directions, discussed in depth in the relevant chapters, are as follows:

- *Methodological improvements*: Measuring aural detectability judgements in future experiments using the novel method is recommended, subject to some revisions. Firstly, if using the method of constant stimuli, the method for identifying the range of stimulus levels should be revised in order to prevent inappropriate sets of stimulus levels being used. The method of adjustment might be effective for achieving this. Secondly, in order to reduce the amount of testing time per condition, an adaptive procedure could be used to more rapidly identify the subject's aural detectability threshold. Should this change to an adaptive procedure be made, a study comparing the results from the method of constant stimuli and the adaptive procedure should be carried out. See Section 7.1 for detailed discussion.
- *Military experience and hearing impairment*: The primary motivation for studying ASA is the potential impact hearing loss might have upon an individual's AFFD. This PhD project in not able to comment on the effects of hearing loss on ASA as all subjects had normal hearing. Future research should therefore investigate the effects of hearing loss on aural detectability judgements, especially the types of hearing loss that are most common in the military and that might affect fitness for duty (i.e. noise-induced hearing loss). Military experience is likely to impact ASA due to previous experience, so this expected effect should also be investigated. Supplementary experiments in this PhD (Appendix A) explored the use of hearing loss simulations for AFFD research. Such simulations could be used in combination with the current VR approach to investigating ASA by using virtual acoustics. This would first require development and validation, but could be a powerful tool for exploring the effect of hearing loss on ASA whilst controlling for military experience.
- *Exploring additional factors*: Whilst discussing the extrinsic and intrinsic factors that might affect ASA, many factors were identified. Using the novel method here with the previously recommended revisions, these factors could be further explored. A high priority factor is the influence of background noise, with the source of the noise either ambient (wind through trees, traffic), near the target (a tank, a generator) or near the subject (vehicle, urban noise). This could be

investigated by additional loudspeakers in an ambisonic array, or through use of virtual acoustics.

- *Field study*: Previous AFFD research has highlighted the importance of measuring the impact of hearing impairment on operational performance in field studies (Sheffield et al., 2016b). These studies have the benefit of high ecological validity, but come at the cost of reduced experimental control. Although efforts have been taken with the current methods for measuring aural detectability judgements, performing a laboratory test will always incur a cost of ecological validity. Therefore, it is recommended that a field study, based on that of Casali et al. (2009) be conducted. Such a study should be mission-based, and involve teams of soldiers conducting a task requiring acoustic stealth. Real-time hearing loss simulators should be used that use masking noise to ensure direct-path sound is inaudible (as determined by Experiment S3, Section A.4). Outcome measures for the task should include GPS tracking, video and audio (calibrated) recording, self-report measures, and objectives measures of aural detectability (e.g. a microphone at the 'enemy' location). Eye tracking information, measured by portable eye-tracking glasses (Mele and Federici, 2012), collected during the task would be highly insightful as compensatory strategies, such as increased reliance on visual cues, may occur with increasing hearing loss.

Appendix A Hearing loss simulation experiments

A.1 Overview

During the early stages of the PhD, the research group was acutely aware of the challenges associated with conducting experiments with military personnel, especially those with hearing impairment, due to two PhD students' experiences (see Section 2.2.4). There are also challenges associated with experimental design for AFFD studies, due to the covariance of age, hearing loss and military experience. To some extent, hearing loss simulation has the potential to overcome these challenges. Other military AFFD researchers have used hearing loss simulators (Sheffield et al., 2017, Semeraro, 2016). However, a limitation of studies using hearing loss simulations is that the results are dependent on the simulation; as such, if the simulation does not give results similar to truly hearing impaired listeners then the results have poor external validity. For hearing loss simulations to be used in order to learn about how hearing loss impacts operational effectiveness and safety, it is highly important that the hearing loss simulation functions accurately.

For AFFD experiments, researchers are interested in both laboratory and field studies. The challenge for hearing loss simulations is therefore: 1) to manipulate audio in a way that produces the same effects as hearing impairment, and 2) do so in a way that permits laboratory and field studies. A number of studies have approached the first challenge by creating simulations inspired by the perceptual deficits of the auditory system, but only a limited number of studies have compared the effects of the hearing loss simulation to listeners with true hearing impairment. The second challenge has received significantly less attention, most likely due to technological limitations.

Due to the limited studies approaching the two aforementioned challenges of hearing loss simulation, three experiments were conducted in order to contribute to this field and further understand the applicability of hearing loss simulation in AFFD research. The first experiment (Section A.2, Experiment S1) measured speech-in-noise (SIN) speech recognition thresholds (SRTs) in a sample of young normal-hearing listeners with hearing loss simulations, and compared the SRTs to a previously collected dataset of SRTs (Athalye, 2010). The second experiment (Section 0, Experiment S2) explored an interesting finding related to the simulation of loudness recruitment. The third experiment (SectionA.4, Experiment S3) was performed in order to assess the feasibility of a real-time version of the hearing loss simulation used in Experiments S1 and S2.

A.2 Experiment S1: A comparison of the effects of real and simulated hearing loss on the perception of speech in noise

A.2.1 Introduction

Hearing loss simulations process sound using algorithms in order to produce the effects of a hearing loss in normal-hearing listeners (Duchnowski, 1989, Villchur, 1974). Most simulations are physiologically inspired, where they attempt to replicate the effects of sensorineural hearing loss, including reduced audibility (Fabry and Van Tasell, 1986), loudness recruitment (Villchur, 1974), reduced frequency selectivity (ter Keurs et al., 1989) and impaired temporal processing (Drullman et al., 1994). Noise-induced hearing loss is the most prevalent type of hearing loss in the military (Humes et al., 2005), so simulating sensorineural hearing loss is of most interest for AFFD research.

A limitation of most studies exploring the effects of hearing loss simulation are small sample sizes (e.g. n = 9 for Moore and Glasberg (1993); n = 4 for Lum and Braida (2000)) and limited comparison to true hearing-impaired listeners using the same methodology, especially when multiple simulations are combined (e.g. simulation of both loudness recruitment and reduced frequency selectivity (Nejime and Moore, 1997)). This experiment aimed to overcome these limitations by testing a larger sample and comparing the results to a dataset obtained using hearing-impaired listeners. The dataset was collected previously at our university as part of the HearCom project (Athalye, 2010); permission to analyse the data was granted (Lutman, 2015). The following three hearing loss simulation algorithms were investigated:

1) A frequency-specific attenuation algorithm to filter the audio signal according to an input audiogram in order to simulate decreased audibility (threshold elevation);

2) A multiband envelope expansion algorithm, designed to simulate the effects of loudness recruitment (Moore et al., 1995, Moore and Glasberg, 1993).

3) A spectral smearing algorithm, designed to simulate the effects of reduced frequency selectivity (Baer and Moore, 1994, Baer and Moore, 1993).

The objectives of the experiment were to measure SIN SRTs using a similar method to that used for the hearing-impaired comparison dataset, under a variety of different processing conditions, in order to achieve the following aims:

1) Explore the effects of and interactions between audibility, loudness recruitment and spectral smearing on the SRT measured in stationary noise.

2) Compare the SRTs measured using the simulations to those measured for hearing-impaired subjects.

A.2.2 Methods

35 normal-hearing (HTLs < 20 dB HL) young (age < 28 years) adult listeners completed the experiment. The monaural SRT was measured via headphones (Sennheiser HDA 200) using the BKB sentences presented in speech-shaped noise. This SRT measurement matched that for the comparison dataset. Only the right ear was tested. A one-up-one-down adaptive procedure was used to measure the SNR (dB) that corresponded to 50% correctly identified key words (SRT). An individual trial was scored correct if 2 or more key words were identified.

The SRT was measured in 10 different processing conditions, shown in Table 8-1. Two audiograms were used: A1 was a 0 dB HL (no attenuation) audiogram, and A2 was equivalent to the average audiogram in the comparison dataset (thresholds: 250 Hz = 27 dB HL; 500 Hz = 27 dB HL; 1000 Hz = 35; 2000 Hz = 52 dB HL; 3000 Hz = 59 dB HL; 4000 Hz = 65 dB HL; 8000 Hz = 65 dB HL). The loudness recruitment expansion factor was based on the audiogram, with higher expansion factors for higher thresholds (the expansion factor was not uniform across frequency). Loudness recruitment was not included in the A1 audiogram conditions as the simulation is only applicable with elevated thresholds. The A1 condition was included in order to assess if the filtering and recombining of the signal affected the SRT. The spectral smearing broadening factor was uniform across frequency. In all conditions the noise was kept at a fixed presentation RMS level and the speech level was varied. To match the comparison dataset, the noise level was fixed at 63 dB SPL for A1 conditions and 73 dB SPL for A2 conditions at the input to the simulation.

Table 8-1 - Processing conditions in Experiment S1.

Condition name	Audiogram	Spectral Smearing	Loudness Recruitment
Norm	N/A	N/A	N/A
A1	0 dB HL flat	Off	Off
A1_S3	0 dB HL flat	Broadening factor x3	Off
A1_S5	0 dB HL flat	Broadening factor x5	Off
A2	Mild-moderate sloping	Off	Off
A2_S3	Mild-moderate sloping	Broadening factor x3	Off
A2_S5	Mild-moderate sloping	Broadening factor x5	Off
A2_LR	Mild-moderate sloping	Off	On
A2_S3_LR	Mild-moderate sloping	Broadening factor x3	On
A2_S5_LR	Mild-moderate sloping	Broadening factor x5	On

Appendix A

Testing was completed using MATLAB. All the hearing loss simulation code for the experiment was written by Dr Jessica Monaghan, and used the Cambridge implementations of envelope expansion (Moore and Glasberg, 1993) and spectral smearing (Baer and Moore, 1993). The adaptive BKB sentence test code was written by Dr Daniel Rowan.

The session started with a familiarisation SRT, equivalent to the Norm condition. Following the familiarisation SRT, the 10 SRTs were measured in a random order. A short break was given, then all 10 SRTs were measured again, in a random order.

This experiment was approved by the University of Southampton Ethics and Research Governance board (Ethics number = 13634).

A.2.3 Results

A repeated-measures ANOVA showed no effect of repeat on the SRT [$F(1,34) = 0.44$, $p > 0.1$], or interaction between repeat and condition [$F(9,306) = 0.907$, $p > 0.1$], so the mean of the two SRTs for each condition was used to represent the SRT for each subject. The mean SRT for each condition for 35 listeners is shown in Figure 8.1.

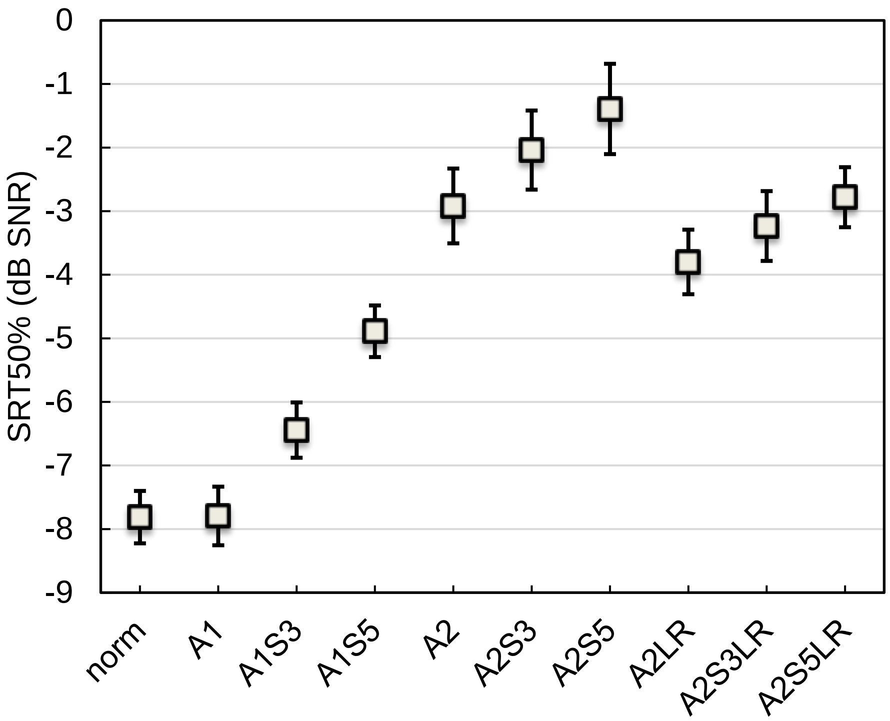

Figure 8.1 - Mean speech recognition threshold (SRT) (n = 35) for each simulation condition in Experiment S1. Error bars indicate the 95% confidence interval of the mean. A lower SRT indicates better performance.

A paired samples t-test showed no difference between norm and A1. Two repeated-measures ANOVA were carried out; one with audiogram (A1 and A2) and spectral smearing (off, S3 and S5) as factors, and one with loudness recruitment (off, LR) and spectral smearing (off, S3, S5) as factors.

For the first ANOVA (audiogram and spectral smearing), there was a significant effect of audiogram, [$F(1,32) = 456.98$, $p < .001$], where A2 conditions gave higher SRTs than A1 conditions. There was a significant effect of spectral smearing, [$F(2,64) = 70.15$, $p < 0.001$], where smearing conditions gave higher SRT scores. A significant interaction between audiogram and spectral smearing was found, [$F(2,64) = 11.76$, $p < 0.001$]. Paired comparisons revealed that the effect of spectral smearing was less for A2 conditions than for A1 conditions.

The second ANOVA (loudness recruitment and spectral smearing) revealed a significant effect of loudness recruitment, [$F(1,34) = 36.36$, $p < 0.001$], and spectral smearing,

[$F(2,64) = 30.22$, $p < 0.001$], and no significant interaction between loudness recruitment and spectral smearing. Pairwise comparisons confirmed that the simulation of loudness recruitment improved the SRT in each spectral smearing condition for A2 conditions.

A comparison between SRTs from hearing loss simulations and the true hearing impaired dataset is shown Figure 8.3. There was a difference between the SRTs for normal-hearing listeners in the comparison dataset and the SRTs measured in the current experiment; in order to compare the effect of hearing loss, the difference between the normal-hearing and hearing-impaired SRT was calculated. This effectively normalised each dataset according to how normal-hearing listeners performed. A higher value indicates a larger detrimental effect of hearing impairment on the SRT.

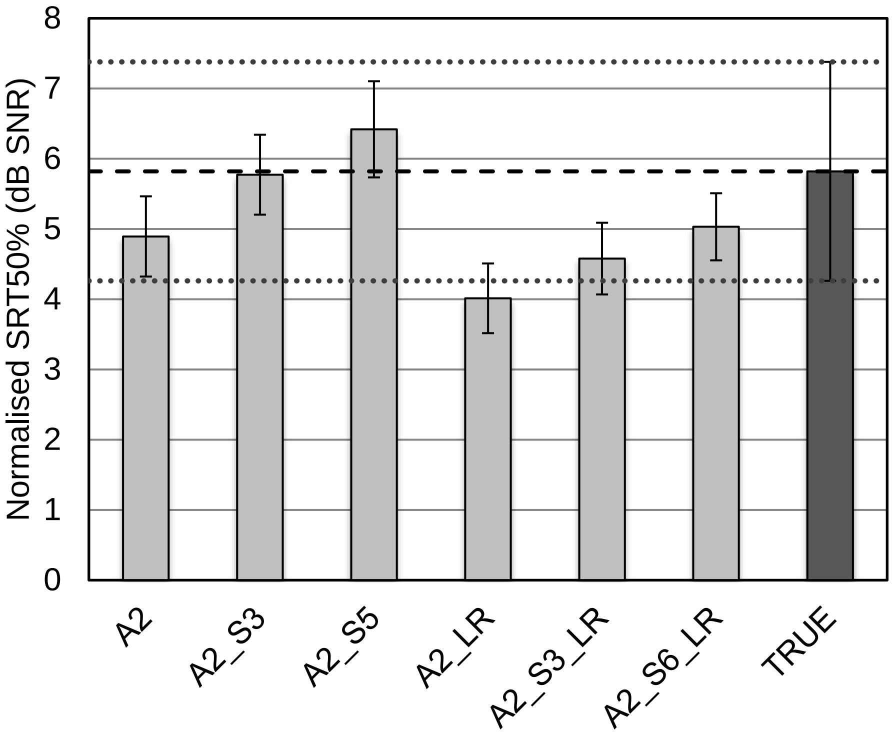

Figure 8.2 - Mean normalised SRT for each simulation condition is shown in light grey and the mean for comparison dataset shown in dark grey on the right. Dashed line shows the mean; dotted lines show the 95% confidence interval of the mean for the true hearing-impaired sample.

A.2.4 Discussion

The results showed that the threshold elevation simulation alone decreased SIN performance. This showed that even though the fixed noise level was 10 dB higher in the A2 conditions, reduced audibility probably affected the scores. It is possible that the upward spread of masking also affected the scores. Spectral smearing decreased SIN performance, with a higher broadening factor leading to poorer performance. The effect was less pronounced in the A2 conditions. This could be related to a reduction in audibility of the target speech; if less of the smeared target speech was audible then the smearing could not have such a detrimental effect. Curiously, the loudness recruitment simulation *improved* the SRT relative to the threshold elevation alone simulation. Based on previous studies, the SRT was expected to be higher (poorer performance) compared to the threshold elevation alone (Moore et al., 1995, Nejime and Moore, 1997). However, the specific set of conditions tested here have not been tested before, making a direct comparison not possible.

In a study of the effects of loudness recruitment simulation on the perception of speech in quiet, a beneficial effect of loudness recruitment was observed for a 50 dB flat loss audiogram relative to an unprocessed condition on the slope of psychometric function (Moore and Glasberg, 1993). The slope was higher for the loudness recruitment condition, suggesting that performance improved at a faster rate for the loudness recruitment simulated speech than for normal speech as the speech level increased. The authors argued that for speech at a given amount above threshold, the loudness would have been greater for the recruitment condition relative to the normal condition, thus increasing the audibility of the signal and therefore performance. A similar effect has been reported in real hearing-impaired listeners with sensorineural losses (recruitment) compared conductive losses (no recruitment) (Gatehouse and Haggard, 1987), where the effect was only present at low sensation levels (when the speech is partially affected by audibility).

The comparison to the true dataset showed that all conditions but one gave similar normalised SRTs to the true hearing-impaired SRTs. The loudness recruitment only condition had a mean normalised SRT outside the 95% confidence interval of the true hearing-impaired normalised SRT. This might have been related to the aforementioned effects of loudness recruitment simulation, but this is unknown. The results showed that the hearing loss simulations negatively the SRT in stationary noise by a similar magnitude to true hearing impairment.

The use of the normalised SRT for comparing SRTs obtained via simulations or true hearing impairment was motivated by the difference in normal-hearing SRTs in each sample. There was approximately 2 dB between the mean SRTs in each sample for the normal-hearing listeners. The reason for this disparity is not clear. Different subjects were used in each sample, so between-subject differences (e.g. differences in hearing thresholds) might explain the disparity. The mean SRT measured in the current study was highly similar to other studies (-7 to -8 dB SNR) employing similar methods (Holmes, 2015), so it is possible that the deviation from the expected results lies with the comparison dataset rather than in the current study. It is recommended that future studies comparing the perception of sound through hearing loss simulations and real hearing impairment collect data using identical methods to minimise any potential disparities, including equipment, length of testing and software.

In conclusion, the hearing loss simulations were shown to affect the SRT in stationary noise by a comparable amount to that observed in a comparison dataset of true hearing-impaired listeners. The hearing loss simulations affected the SRT as expected in most conditions, but the loudness recruitment simulation improved the SRT, at least in the conditions tested here. Further work is required to explore this effect.

A.3 Experiment S2: Further exploration of the effects of loudness recruitment on the perception of speech in noise

A.3.1 Introduction

Following the results of Experiment S1, it was decided that an additional experiment was required to further explore the apparent beneficial effect of loudness recruitment simulation on the SRT measured in stationary noise relative to threshold elevation simulation. The exact reason for this beneficial effect was not clear, so a set of conditions was designed to explore two hypotheses.

In Experiment S1, the speech was likely to have been affected by audibility in the hearing-impaired audiogram conditions due to the effect of the threshold elevation simulation alone. The improvement in SRT for loudness recruitment conditions relative to threshold elevation could be explained by increased audibility of the target speech as a result of the loudness recruitment simulation. If this is the case, the beneficial effect should only be present in conditions where the audibility of speech is affected, i.e. low sensation levels. Therefore, four conditions were used to test this hypothesis, seen in Table 8-2, conditions 1 – 4. A flat 40 dB HL audiogram was used for all 4 conditions. Two presentation levels were used, one close to threshold and therefore leading to reduced audibility (64 dB SPL) and one well above threshold (79 dB SPL) that was less affected by audibility. For each presentation level, the audio was processed using either the threshold elevation or loudness recruitment simulation.

An additional four conditions were tested as part of this experiment in order to test another hypothesis for the beneficial effect of loudness recruitment simulation relative to threshold elevation alone. However, after the experiment it became apparent that the code required to test this hypothesis was problematic; therefore, the analysis for these condition is not reported.

A.3.2 Methods

20 normal-hearing young listeners took part in the experiment. SRTs were measured using the same procedure as Experiment S1 (see Section A.2.2 for full details). The input audiogram for the hearing loss simulation was a 40 dB HL flat loss. For the loudness recruitment simulation, this results in an expansion factor of 1.67. Each condition was as described in the introduction and shown in Table 8-2. Note, the presentation level is equivalent to the level at the input to the simulation. Each SRT was measured once due to the good test-retest reliability shown in Experiment S1. The MATLAB code was written by Dr Jessica Monaghan and the same BKB sentence test, written by Dr Daniel Rowan, was used.

For the purpose of transparency of reporting what each subject experienced during the experimental session, the additional four conditions that were later found to be problematic used the an identical SRT procedure as for Experiment S1, but had the expansion factor manipulated but the output level fixed at 63 dB SPL. Each subject completed 8 SRTs measurements following a familiarisation SRT measurement.

This experiment was approved by the University of Southampton Ethics and Research Governance board (Ethics number = 13634.A1).

Table 8-2 - Conditions for Experiment S2. TE represents threshold elevation, LR represents loudness recruitment.

Condition number	Condition name	Processing condition
1	64_TE	Presentation level = 64 dB SPL, threshold elevation
2	64_LR	Presentation level = 64 dB SPL, loudness recruitment
3	79_TE	Presentation level = 79 dB SPL, threshold elevation
4	79_LR	Presentation level = 79 dB SPL, loudness recruitment

A.3.3 Results

The results are shown in Figure 8.3.

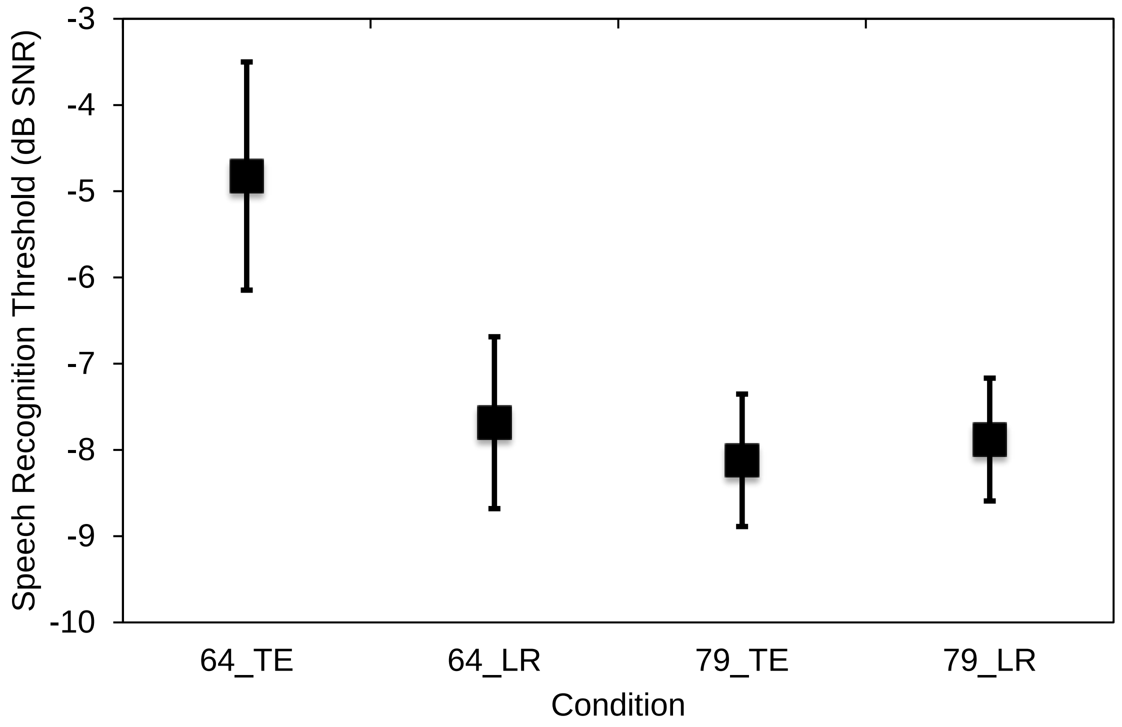

Figure 8.3 Experiment S2 results. Error bars indicate the 95% confidence interval of the mean.

A repeated-measures ANOVA with presentation level and processing condition as factors showed a significant effect of presentation level [$F(1,16) = 19.10$, $p < 0.001$], a significant effect of processing condition [$F(1,16) = 33.61$, $p < 0.001$], and a significant interaction between input level and processing condition [$F(1,16) = 26.23$, $p < 0.001$].

Pairwise comparisons showed a significant difference between processing conditions for the 64 dB SPL presentation level [$t(19) = 7.95$, $p < 0.001$], where the loudness recruitment SRT (64_LR) was lower than the threshold elevation condition (64_TE), indicating a lower SRT for loudness recruitment for this presentation level of 2.86 dB SNR (2.11 to 3.61 dB SNR 95% confidence interval) relative to threshold elevation. No significant difference was found between processing conditions for the 79 dB input level [$t(19) = -0.61$, $p = 0.55$].

A.3.4 Discussion

This experiment explored the apparent beneficial effect of loudness recruitment. The condition included tested the hypothesis that expanded speech increases the loudness of the speech relative to unexpanded, therefore in conditions where the audibility of the target speech is affected by elevated thresholds, the expanded target speech performance in SIN is expected at low sensation levels where audibility of speech is a factor, but not at high levels, as no changes in audibility occur.

The results support the hypothesis that the loudness recruitment simulation has a beneficial effect relative to threshold elevation alone at low sensation levels due to increased audibility of the target speech, as seen in the improvement in SRT at the low presentation level but not at the high level. The increase in audibility of the target speech means the subject can achieve a lower SNR (better performance), because a higher noise level can be tolerated. The results of these experiments suggest that the envelope expansion alone neither improves or degrades SIN performance when audibility is not compromised, but it does improve the SRT when the speech is at low sensation levels and the audibility is compromised.

Experiment S1 showed that the hearing loss simulations used gave similar speech in noise perception to a real hearing-impaired sample. The apparent beneficial effect of loudness recruitment relative to threshold elevation alone was shown in Experiment S2 to only be present when the target speech is at low sensation levels. Therefore, for AFFD studies requiring hearing loss simulation it is recommended that the simulation includes threshold elevation, loudness recruitment and spectral smearing. Future work that could increase the confidence in the validity of hearing loss simulations for AFFD studies should compare speech perception in listening situations more ecologically valid to those experienced in occupational settings.

A.4 Experiment S3: Feasibility study of a real-time hearing loss simulator

A.4.1 Introduction

The need for a hearing loss simulation that is able process sound in real-time and is sufficiently portable to enable AFFD field studies was highlighted previously (Section A.1). This section describes the evaluation of a real-time hearing loss simulation mobile application that was developed by Dr Jessica Monaghan at the University of Southampton.

The mobile application uses the microphone and audio output of a smartphone (see Figure 8.6 for a photo of the device). The application simulates threshold elevation, loudness recruitment and spectral smearing. Due to the encouraging results of Experiments S1 and S2, this device is a potentially powerful tool for outdoor AFFD studies in order to measure the impact of hearing loss with greater ecological validity.

There are two potential limitations to the device. Firstly, if the latency of the audio output is too high then the user will probably not tolerate wearing the device. Decrements in the performance of a soldier might be due to increased cognitive load associated with the audio delay, rather than the hearing loss itself. Therefore, the latency of the device must be sufficiently low. Studies investigating tolerable delay in hearing aids suggest that delays of up to 20 msec are tolerable (Stone and Moore, 1999, Goehring et al., 2018). Secondly, for the hearing loss simulation to be valid, the audio that the user experiences must only be that of the simulation. As such, the direct path (perception of the external sound source *not* via the device) must be inaudible, or at least less intense than the output of the simulation device. The real-ear attenuation at threshold (REAT) must therefore be sufficiently high to allow hearing loss to be investigated. In order to maximise the chances that the REAT was sufficiently high, the device used insert headphones with ear muffs with the highest attenuation levels on top.

The aim of this study was to assess the latency and the REAT of the device in order to assess the feasibility that the real-time hearing loss simulation device could be used for AFFD studies.

A.4.2 Methods

The latency of the device was assessed using a KEMAR mannequin. One ear was left open and the other had the signal delivered via the hearing loss simulation device. A stereo recording was made of intense transient sounds from 1 m away at 0° azimuth. The latency was calculated by finding the difference between peaks of the left and right channels. This was repeated for various hearing loss simulation settings. A photo of the hearing loss simulation device in-situ on the KEMAR mannequin is shown in Figure 8.6.

Figure 8.4 - Real-time hearing loss simulation device worn by the KEMAR mannequin.

The average REAT was measured behaviourally for 16 young healthy normal-hearing listeners. The method was based on the British standard for the subjective measurement of attenuation in hearing protection devices (BS, 1993), with a small number of adaptations. The REAT was found by measuring sound-field hearing threshold levels (HTLs) in an anechoic chamber, and calculating the difference between the thresholds with an open ear and with the device in-situ. The test signals were frequency modulated tones (warble) at 125, 250, 500, 1000, 2000, 4000 and 8000 Hz. Thresholds at each frequency were measured twice in each listening condition. Thresholds were measured for four conditions: open ears, ear muffs only, inserts only, and ear muffs and inserts. The hearing loss simulation device was switched off in all conditions. The threshold measuring procedure was based on the British standard for pure tone audiometry (BS, 2009).

Sounds were generated by MATLAB and a Genelec loudspeaker placed 1.5 m directly ahead of the listener. The code was written by Dr Mark Fletcher (the 'Uberometer'). This experiment was approved by the University of Southampton Ethics and Research Governance board (Ethics number = 20032).

A.4.3 Results

The latency of the device with each setting is shown in

Table 8-3. The average REAT for each condition at each frequency is shown in Figure 8.7.

Device condition	Latency (msec)
Off	0
Normal	18
Threshold elevation	22
Loudness recruitment	23

Table 8-3 - Latency of real-time hearing loss simulation device with various simulation settings. Note, ‘All’ describes the device condition that simulated loudness recruitment and spectral smearing.

Device condition	Latency (msec)
Off	0
Normal	18
Threshold elevation	22
Loudness recruitment	23
All	24

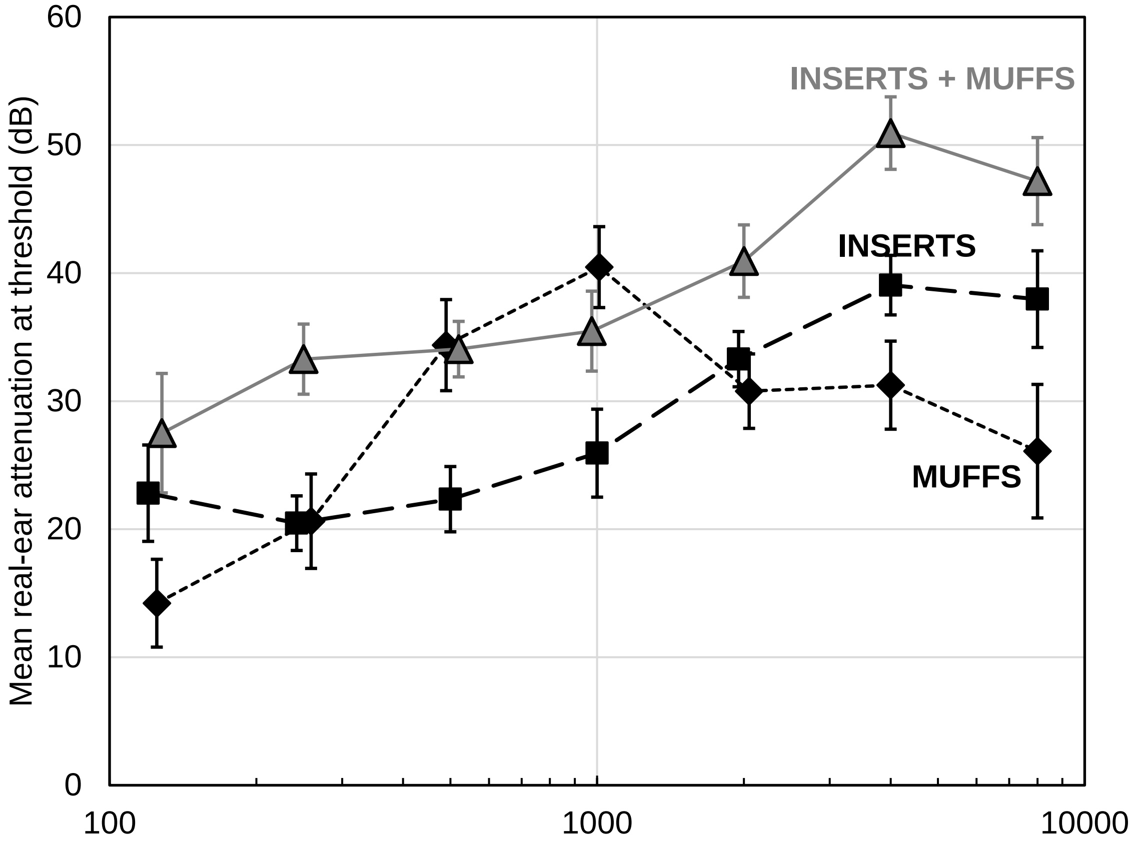

Figure 8.5 - Average real-ear attenuation at threshold for the hearing loss simulation device. Error bars indicate 95% confidence intervals of the mean.

A.4.4 Discussion

The latency of the device was at the threshold of the maximum permissible. The latency is therefore not considered a problem, though future developments should consider reducing the latency.

The maximum REAT shows the maximum hearing loss that can be simulated. If the hearing loss to be simulated is greater than the REAT, the user will perceive sound even though the iPhone output is silent. Therefore, the maximum hearing loss that can be simulated using the device is a mild low-frequency hearing loss of up to 30 dB HL up to 1000 Hz, rising to 40 dB HL above 1000 Hz.

For AFFD studies, a range of hearing loss configurations are of interest. In particular, a hearing loss in the H3 category (see Section 2.2.2 for details), equivalent to a moderate hearing loss, is required, because H3 hearing loss is usually considered not fit for duty. Therefore, because the real-time hearing loss simulation device is unable to attenuate the direct path of sound required for simulating a H3 hearing loss, the device in its current form requires further development.

Two alternative approaches could overcome the limitations of the device identified as a result of this study. Active sound reduction technology could potentially increase the REAT. This involves a microphone placed at the ear and presenting an inverted version of the signal to the ear, which cancels out the sound to a certain extent. Active sound reduction on its own typically has less overall REAT than passive attenuation and is not very effective in the high frequencies (Rudzyn and Fisher, 2012), but a combination of in-ear active headphones with passive ear muffs might yield greater REAT.

Another approach is to present noise, spectrally shaped to match the desired audiogram, to the subject in order to mask the direct path of sound. Whilst this has the advantage of temporarily raising the thresholds of the user, it comes at the cost of reduced hearing loss simulation validity. This approach was used by Sheffield et al. (2017), who used a real-time hearing loss simulation device with masking noise and loudness recruitment simulation. Devices such as these are lacking validation, so future studies should consider conducting a study, such as Experiment S1, in order to increase confidence in the ability of the device to accurately simulate hearing loss.

In conclusion, the device tested in the current study, in its current form, was not considered appropriate for use in AFFD studies due to limited REAT. Given the restrictions on the hearing loss magnitudes that can be simulated due to limited REAT, it is likely that real-time hearing loss simulation device will need to include masking. Further studies are required to validate this approach.

Appendix B MATLAB code: Outdoor sound propagation model

The following MATLAB code was written to attenuate a sound as if it had propagated from source to a receiver location, given the source-to-reference distance, source-to-receiver distance, reference height and receiver height.

```
%% Sound propagation model
% Written by MB
% October 2018

% Takes audio signal at reference location (r1) and attenuates
signal
% according to ISO9613-2 for outdoor propagation to target
location (r2).
% Takes into account:
% - Geometric spreading (uses source-to-reference and source-to-
target
% distances)
% - Atmospheric absorption (uses distance and absorption
coefficients)
% - Ground effect (uses ground factors G, source height and
receiver height

% Steps in model:
% 1. Split into octave bands
% 2. Apply octave band specific attention according to ISO9613-2
% 3. Recombine channels

% Inputs:
% - signal_in = signal to be attenuated. 1 x n vector, already
read in (not
% a .wav file!)
% - r1 = source-to-reference distance (m)
% - r2 = source-to-target distance (m)
% - hs = source height (m)
% - hr = receiver height (m)
% - fs = sampling frequency


function [PostModelSignal,rms_atten] =
PropagationModel(signal_in,r1,r2,hs,hr,fs)
% Preliminary stuff
sig = signal_in;
CFs = [63 125 250 500 1000 2000 4000 8000 16000]; % Centre
frequencies
xo = [88 177 355 710 1420 2840 5680 11360]; % Crossover
frequencies
BP = zeros(length(CFs),length(sig)); % matrix that holds the
filtered signal
BP_Post = zeros(length(CFs),length(sig)); % matrix that holds the
post-attenuation signal
```

```
samples = 1:length(sig);
t = samples/(fs);

% Split into channels. Splits it into channels as defined by
crossover
% frequencies (xo).
for i = 1:length(xo)
    if i == 1
        %first
        [sHigh,sLow] = xoFilt(sig,xo(i),fs);
        BP(i,:)=sLow;
    elseif (i > 1) && (i < length(xo))
        %middle
        [sHigh,sLow] = xoFilt(sHigh,xo(i),fs);
        BP(i,:)=sLow;
    else
        %final
        [sHigh,sLow] = xoFilt(sHigh,xo(i),fs);
        BP(i,:)=sLow;
        BP(i+1,:)=sHigh;
    end
end

% Sound propagation model
for i =1:length(BP(:,1))
    tempSig = BP(i,:);
    tempSig = GeometricAttenuation(tempSig,r1,r2);
    tempSig = AtmosphericAbsorptionAttenuation(tempSig,r2,i);
    tempSig = GroundEffectAttenuation(tempSig,r2,i,hs,hr,1);
    BP_Post(i,:)=tempSig;
end

% Recombine channels (sum each sample)
attenSig = zeros(1,length(BP_Post));
for i = 1:length(BP_Post)
   attenSig(i) = sum(BP_Post(:,i));
end
attenSig = attenSig';
rms_atten = (20*log10(rms(sig)/2e-5))-(20*log10(rms(attenSig)/2e-
5));
PostModelSignal = attenSig;

end

%% Crossover function
% Used in loop to cascade filters to seperate original signal into
octave
% channels.
% Variables:
% sig_high = high passed signal
% sig_low = low passed signal
% sig_in = signal to filter
% xo_freq = crossover frequency. For example, the crossover
frequency
% between channels with centre frequencies 63 and 125 is 88.
% fs = sampling frequency
function [sig_high,sig_low] = xoFilt(sig_in,xo_freq,fs)
sig_low = LRfilt(sig_in,0,xo_freq,fs);
sig_high = LRfilt(sig_in,1,xo_freq,fs);
```

```
end

%% Filter function.
% Step 1: designs filter. Uses order 2 butterworth filter.
% Step 2: does forward and backward filter. This ensures zero-
phase and
% effectively doubles up to a 6 dB attenuation at crossover
frequency. This
% has the effect of a Linkwitz-Riley filter, such that when
recombined, a
% flat response of the filterbank is given.

% Variables
% sig: signal to be filtered
% lphp: determines high pass or low pass. 0 = lp, 1 = hp.
% freq_cutoff: cutoff frequency
% fs: sampling frequency
function out_filt = LRfilt(sig,lphp,freq_cutoff,fs)
if lphp == 0
    [b,a] = butter(2,hz2normFreq(freq_cutoff,fs),'low');
elseif lphp == 1
    [b,a] = butter(2,hz2normFreq(freq_cutoff,fs),'high');
end
out_filt = filtfilt(b,a,sig);
end

%% Converts frequency (Hz) to normalised frequency (between 0 and
1, where 1
% is nyquist frequency). Used in butterworth filter design.
function normFreq = hz2normFreq(freqHz,fs)
normFreq = freqHz/(fs/2);
end

%% Geometric absorpotion function (20log10(r2/r1)).
% sig_in: signal to be attenuated.
% r1 = reference distance from source (m).
% r2 = target distance from source (m).
% sig_out = attenuated signal.
function sig_out = GeometricAttenuation(sig_in,r1,r2)
sig_out = sig_in*db2mag((20*log10(r1/r2)));
end

%% Atmospheric absorption formula (ISO9613-2)
% sig_in = signal to be attenuated (octave wide channel)
% r2 = distance from source (m).
% channel = centre frequency channel (used to determine which
absorbtion
% coefficient to use).
% sig_out = attenuated signal.
function sig_out =
AtmosphericAbsorptionAttenuation(sig_in,r2,channel)
    a = [0.1 0.3 1.1 2.8 5 9 22.9 76.6 229.2]; %absorption
coefficients from ISO9613-2 (in dB/km)
    sig_out = sig_in/db2mag((a(channel)*r2)/1000);
end

%% Ground effect (ISO9613-2)
% sig_in = signal to be attenuated (octave wide channel)
% r2 = distance from source (m).
```

```
% channel = centre frequency channel (used to determine which
absorbtion
% coefficient to use).
% sig_out = attenuated signal.
% hs = source height (m).
% hr = receiver height (m).
% G = ground factor (1 for grass).
function sig_out =
GroundEffectAttenuation(sig_in,r2,channel,hs,hr,G)
%calculate q for Am
if r2 <= 30*(hs+hr)
    q = 0;
else
    q = 1-((30*(hs+hr))/r2);
end

switch channel
    case 1 % 63 Hz CF octave band
        As = -1.5;
        Ar = -1.5;
        Am = (q^2)*-3;
    case 2 % 125 Hz CF octave band
        asFactor = 1.5 + ((3*exp((-0.12*(hs-5)^2)))*(1-(exp((-
1*r2)/50)))) + 5.7*(exp(-0.09*hs^2)*(1-exp(-2.8*10^(-6*r2^2))));
        As = (-1.5)+(G*asFactor);
        arFactor = 1.5 + ((3*exp((-0.12*(hr-5)^2)))*(1-(exp((-
1*r2)/50)))) + 5.7*(exp(-0.09*hr^2)*(1-exp(-2.8*10^(-6*r2^2))));
        Ar = (-1.5)+(G*arFactor);
        Am = -3*q*(1-G);
    case 3 % 250 Hz CF octave band
        bsFactor = 1.5 + 8.6*(exp(-0.09*hs^2))*(1-exp((-
1*r2)/50));
        As = (-1.5) + (G*bsFactor);
        brFactor = 1.5 + 8.6*(exp(-0.09*hr^2))*(1-exp((-
1*r2)/50));
        Ar = (-1.5) + (G*brFactor);
        Am = -3*q*(1-G);
    case 4 % 500 Hz CF octave band
        csFactor = 1.5 + 14.0*(exp(-0.46*hs^2))*(1-exp((-
1*r2)/50));
        As = (-1.5) + (G*csFactor);
        crFactor = 1.5 + 14.0*(exp(-0.46*hr^2))*(1-exp((-
1*r2)/50));
        Ar = (-1.5) + (G*crFactor);
        Am = -3*q*(1-G);
    case 5 % 1000 Hz CF octave band
        dsFactor = 1.5 + 5.0*(exp(-0.9*hs^2))*(1-exp((-1*r2)/50));
        As = (-1.5) + (G*dsFactor);
        drFactor = 1.5 + 5.0*(exp(-0.9*hr^2))*(1-exp((-1*r2)/50));
        Ar = (-1.5) + (G*drFactor);
        Am = -3*q*(1-G);
    case 6 % 2000 Hz CF octave band
        As = -1.5*(1-G);
        Ar = -1.5*(1-G);
        Am = -3*q*(1-G);
    case 7 % 4000 Hz CF octave band
        As = -1.5*(1-G);
        Ar = -1.5*(1-G);
```

```
            Am = -3*q*(1-G);
        case 8 % 8000 Hz CF octave band
            As = -1.5*(1-G);
            Ar = -1.5*(1-G);
            Am = -3*q*(1-G);
        case 9 % 16000 Hz CF octave band
            As = -1.5*(1-G);
            Ar = -1.5*(1-G);
            Am = -3*q*(1-G);
    end
    %total attenuation (ISO9613-2) is Agr.
    Agr = As + Ar + Am;
    sig_out = sig_in/db2mag(Agr);
    end
```

Appendix C Virtual reality environment development

The development of the virtual reality environment used in Experiment 2 and 3 is reported in this appendix. The environment to be developed was a replica of an area within Southampton Common. The first stage involved gathering information about the dimensions and visual features of the environment. This was achieved by taking detailed photos around the test site, mostly at the location where the subject would be positioned during the experiments. In order to ensure the dimensions of the field were accurately preserved, a Google Earth satellite image was used. Using the scale of the image, it was possible to recreate the dimensions of boundaries of the field (e.g. where the tree-line edges of the field were). This was done by overlaying a grid with the same scale to the satellite image, and building the virtual environment based on each square within the grid.

The virtual environment was built using Unity 5.5 (Technologies, 2017). Virtual environments are created by using the various native terrain building features available in Unity. Unity has units of space within virtual environments, and this directly associates to the mapping of space in virtual reality. One unit of Unity space is equivalent to 1 metre. The environment was therefore built with this scale (1 unit of Unity space per 1 metre of real-world space).

Of most importance for the environment created here was the accurate representation of trees and the ground, as these involve important depth perception cues (texture gradient, relative size). A variety of trees were used in order to represent the variety in tree species found in the real environment. These were placed manually in the virtual environment. The photos of the test site were used to ensure the height of the trees were represented appropriately in the virtual environment. Constructing virtual grass presented computational and ecological validity challenges. Individual blades of grass can be built in the virtual environment to give a compelling representation of real grass; however, this is computationally expensive, especially for an area of open field of the size used here. Upon doing this, the frame rate decreased significantly and led to unacceptable experiences in VR. Upon reviewing many forums, specialist websites and blogs, this problem appeared common and is solved by using textures rather than three dimensional objects. This involves ‘painting’ the ground with a colour and texture that is representative of grass. This reduces the ecological validity of the grass in the scene, but it allows good frame rates to be preserved. For the experiment, it was considered highly important that texture cues were preserved in the virtual environment; this was still achieved by the method adopted here, as the relative density of detail in the grass decreases as the distance increases, as in reality.

Implementing the virtual environment with virtual reality headsets is simple due to the software packages developed by Unity and Oculus Rift. In order to control the virtual environment independently from the view of the user (the view from the VR headset), the environment was developed as an asymmetric game. This means two independent views of the same virtual environment are produced. For example, two player computer games split the screen so that each player experiences the environment from independent perspectives. This same approach was used here, with player one's view (the subject) being cast onto the VR headset and player two (the experimenter) being cast on to the laptop screen. The view of, and the controls for, the virtual environment for the experimenter are shown in Figure 8.6. The subject is represented by the sphere ('My head').

Figure 8.6 - Screenshot of virtual environment in Experiments 2 and 3.

Appendix D Experiment 2 stimulus presentation levels

Table 8-4 – Mean stimulus levels used in experiment 2. See stimulus and piloting section in Chapter 4, and methods section of Experiment 2 in Chapter 5 for full details.

Stimulus	Target distance (m)	Profile	Levels (dB SPL)					
			1	*2*	*3*	*4*	*5*	*6*
Pine cone	25	Familiarisation	-2.5	7.5	17.5	27.5	37.5	47.5
		A	-2.5	6.5	10.5	14.5	18.5	27.5
		B	7.5	16.5	20.5	24.5	28.5	37.5
		C	17.5	26.5	30.5	34.5	38.5	47.5
		D	27.5	36.5	40.5	44.5	48.5	57.5
	50	Familiarisation	4.1	14.1	24.1	34.1	44.1	54.1
		A	4.1	13.1	17.1	21.1	25.1	34.1
		B	14.1	23.1	27.1	31.1	35.1	44.1
		C	24.1	33.1	37.1	41.1	45.1	54.1
		D	34.1	43.1	47.1	51.1	55.1	64.1
	100	Familiarisation	11.1	21.1	31.1	41.1	51.1	61.1
		A	11.1	20.1	24.1	28.1	32.1	41.1
		B	21.1	30.1	34.1	38.1	42.1	51.1
		C	31.1	40.1	44.1	48.1	52.1	61.1
		D	41.1	50.1	54.1	58.1	62.1	71.1
Noise	25	Familiarisation	14.9	24.9	34.9	44.9	54.9	64.9
		A	14.9	23.9	27.9	31.9	35.9	44.9
		B	24.9	33.9	37.9	41.9	45.9	54.9
		C	34.9	43.9	47.9	51.9	55.9	64.9
		D	44.9	53.9	57.9	61.9	65.9	74.9
	50	Familiarisation	21.4	31.4	41.4	51.4	61.4	71.4
		A	21.4	30.4	34.4	38.4	42.4	51.4
		B	31.4	40.4	44.4	48.4	52.4	61.4
		C	41.4	50.4	54.4	58.4	62.4	71.4
		D	51.4	60.4	64.4	68.4	72.4	81.4
	100	Familiarisation	28.4	38.4	48.4	58.4	68.4	78.4
		A	28.4	37.4	41.4	45.4	49.4	58.4
		B	38.4	47.4	51.4	55.4	59.4	68.4
		C	48.4	57.4	61.4	65.4	69.4	78.4
		D	58.4	67.4	71.4	75.4	79.4	88.4
Whisper	25	Familiarisation	7.3	17.3	27.3	37.3	47.3	57.3
		A	7.3	16.3	20.3	24.3	28.3	37.3
		B	17.3	26.3	30.3	34.3	38.3	47.3
		C	27.3	36.3	40.3	44.3	48.3	57.3
		D	37.3	46.3	50.3	54.3	58.3	67.3
	50	Familiarisation	13.7	23.7	33.7	43.7	53.7	63.7
		A	13.7	22.7	26.7	30.7	34.7	43.7
		B	23.7	32.7	36.7	40.7	44.7	53.7
		C	33.7	42.7	46.7	50.7	54.7	63.7
		D	43.7	52.7	56.7	60.7	64.7	73.7
	100	Familiarisation	20.5	30.5	40.5	50.5	60.5	70.5
		A	20.5	29.5	33.5	37.5	41.5	50.5
		B	30.5	39.5	43.5	47.5	51.5	60.5
		C	40.5	49.5	53.5	57.5	61.5	70.5
		D	50.5	59.5	63.5	67.5	71.5	80.5

Appendix E Complimentary figures and tables for experiment 2

E.1 Overview

The figures and tables displayed in this appendix show the data for each participant in Experiment 2. Figures are displayed for each participant, for each condition and for each session. As a function of RMS level experienced by the subject (dB SPL), each figure shows:

- The proportion of 'yes' responses (p(yes))
- The fitted psychometric function.

A legend explaining what the lines and points represent is given in Table 8-5. The tables show the location, slope, error associated with the location and slope estimates, deviation and its associated *p*-value (pdev) for each condition for each participant.

For detailed discussion of the methodology used in Experiment 2, see Chapter 4. For a summarised version of the methods, see section 0. For methodology of psychometric function fitting, see section 5.4.3. Note, the lapse rate and guess rate were fixed to 0.01 and 0, respectively.

Table 8-5 - Legend for figures included in complimentary figures for Experiment 2

Condition	**Feature**	**Description**
25 m target distance, session 1	Blue circle, blue dotted line	Raw results
	Blue solid line	Fitted psychometric function
25 m target distance, session 2	Red x, red dotted line	Raw results
	Red solid line	Fitted psychometric function
50 m target distance, session 1	Green square, green dotted line	Raw results
	Green solid line	Fitted psychometric function
50 m target distance, session 2	Pink diamond, pink dotted line	Raw results
	Pink solid line	Fitted psychometric function
100 m target distance, session 1	Turquoise right-pointed triangle, turquoise dotted line	Raw results
	Turquoise solid line	Fitted psychometric function
100 m target distance, session 2	Black left-pointing triangle, black dotted line	Raw results
	Black solid line	Fitted psychometric function

E.2 Figures and tables

8.2.1.1 *Participant 1*

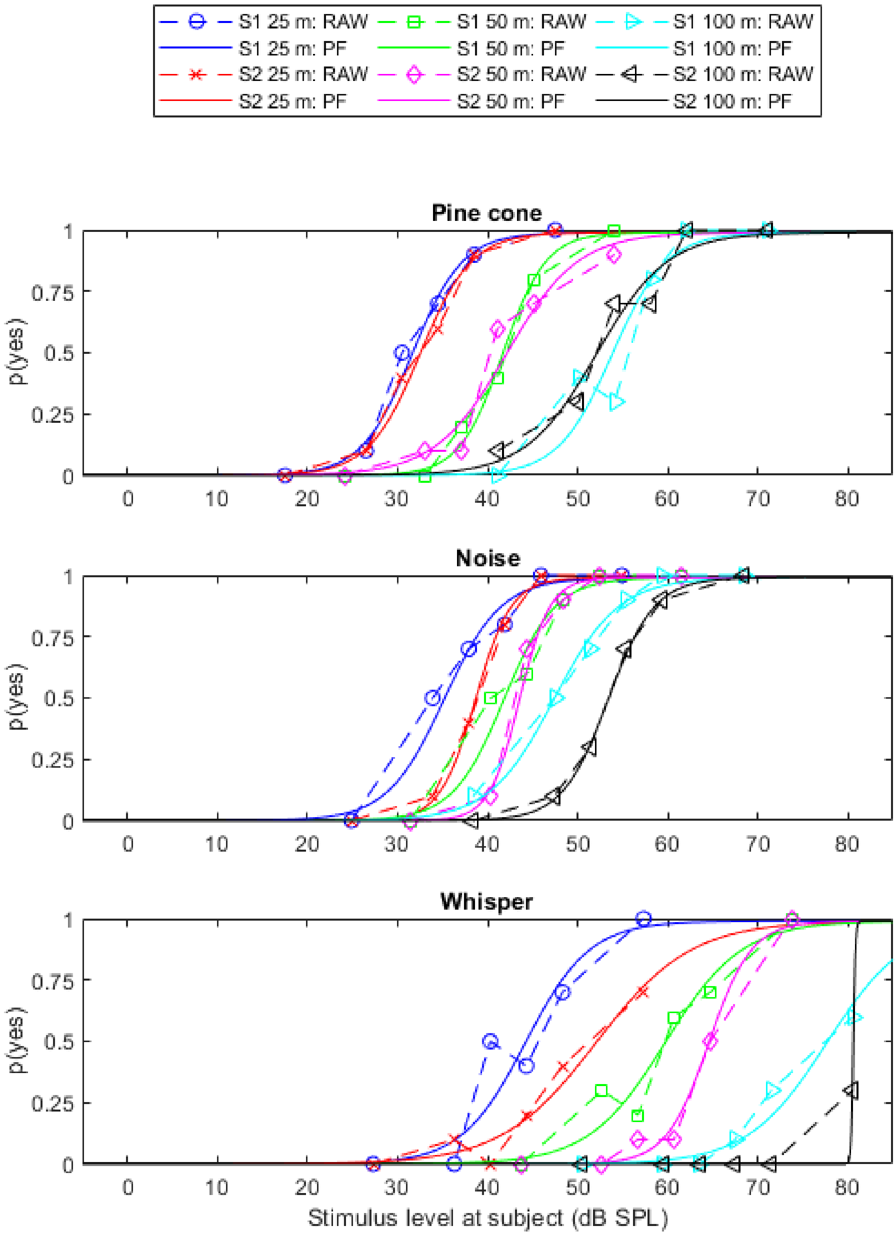

Figure 8.7 - Experiment 2. Individual results and psychometric functions for participant number 1. See Appendix E overview for full details and legend.

Table 8-6 - Experiment 2. Individual psychometric function parameter estimates for participant number 1. See Appendix E overview for full details.

Stimulus	Target distance (m)	Session	Location (dB SPL)	Location SE (dB)	Slope (β)	Slope SE (β)	Deviation	pDev
Pine cone	25	1	31.62	1.13	0.36	0.23	0.85	0.83
		2	32.50	1.13	0.35	0.18	0.63	0.91
	50	1	41.67	1.12	0.42	0.38	1.14	0.72
		2	42.08	1.37	0.25	0.10	2.81	0.59
	100	1	53.84	1.17	0.33	0.16	5.16	0.17
		2	52.08	1.54	0.25	0.10	3.32	0.50
Noise	25	1	35.18	1.37	0.32	0.16	2.62	0.49
		2	38.66	0.87	0.51	0.48	0.67	0.73
	50	1	41.82	1.15	0.35	0.19	2.21	0.48
		2	43.55	0.83	0.59	0.56	0.83	0.48
	100	1	47.37	1.46	0.28	0.13	0.98	0.87
		2	53.40	1.00	0.39	0.24	0.17	0.96
Whisper	25	1	44.20	1.29	0.28	0.11	6.24	0.15
		2	52.04	2.74	0.19	0.06	2.96	0.50
	50	1	59.71	1.49	0.24	0.08	3.50	0.48
		2	64.56	1.14	0.42	1.04	1.75	0.44
	100	1	77.66	2.71	0.23	0.47	1.88	0.61
		2	80.60	0.74	8.50	0.70	0.00	0.74

8.2.1.2 *Participant 2*

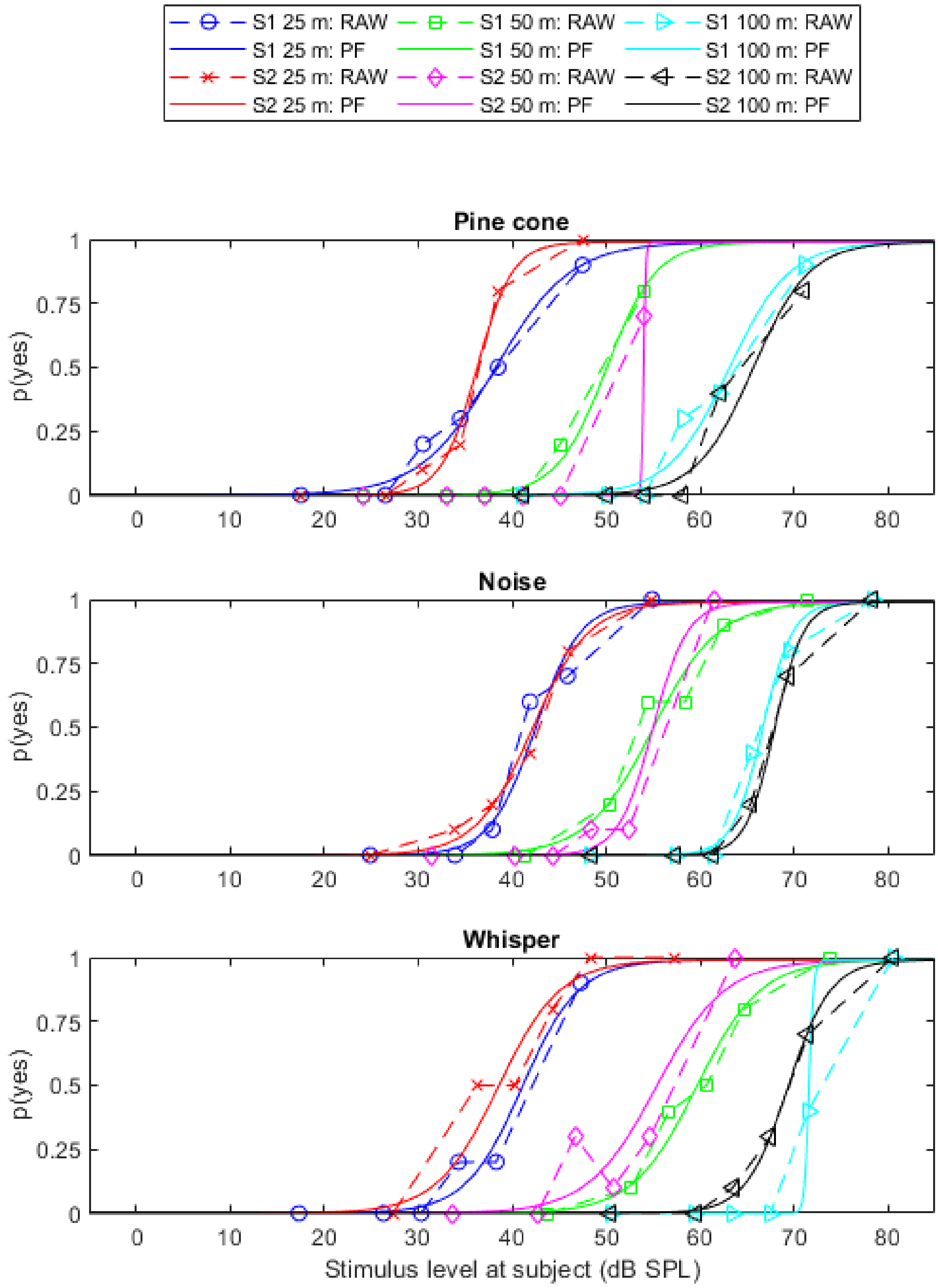

Figure 8.8 - Experiment 2. Individual results and psychometric functions for participant number 2. See Appendix E overview for full details and legend.

Table 8-7 - Experiment 2. Individual psychometric function parameter estimates for participant number 2. See Appendix E overview for full details.

Stimulus	Target distance (m)	Session	Location (dB SPL)	Location SE (dB)	Slope (β)	Slope SE (β)	Deviation	pDev
Pine cone	25	1	38.19	1.70	0.25	0.21	1.61	0.79
		2	36.25	0.99	0.53	0.76	1.12	0.39
	50	1	50.03	2.23	0.38	1.64	1.22	0.32
		2	54.00	0.66	8.51	1.02	0.00	0.57
	100	1	63.09	1.47	0.31	0.72	2.59	0.45
		2	65.62	1.91	0.33	1.13	3.71	0.21
Noise	25	1	42.47	1.06	0.42	0.30	2.43	0.39
		2	42.16	1.14	0.33	0.27	0.87	0.86
	50	1	54.99	1.27	0.29	0.11	2.16	0.61
		2	54.96	1.93	0.55	2.41	2.58	0.09
	100	1	66.66	0.83	0.62	0.78	1.31	0.28
		2	67.93	0.94	0.64	1.05	0.45	0.41
Whisper	25	1	41.09	1.77	0.33	1.07	2.09	0.46
		2	38.53	1.31	0.30	0.11	3.76	0.35
	50	1	59.63	1.23	0.29	0.10	1.27	0.82
		2	55.35	1.55	0.27	0.29	8.38	0.06
	100	1	71.60	0.54	4.13	0.46	0.00	0.41
		2	69.37	1.09	0.44	0.49	0.55	0.77

8.2.1.3 *Participant 3*

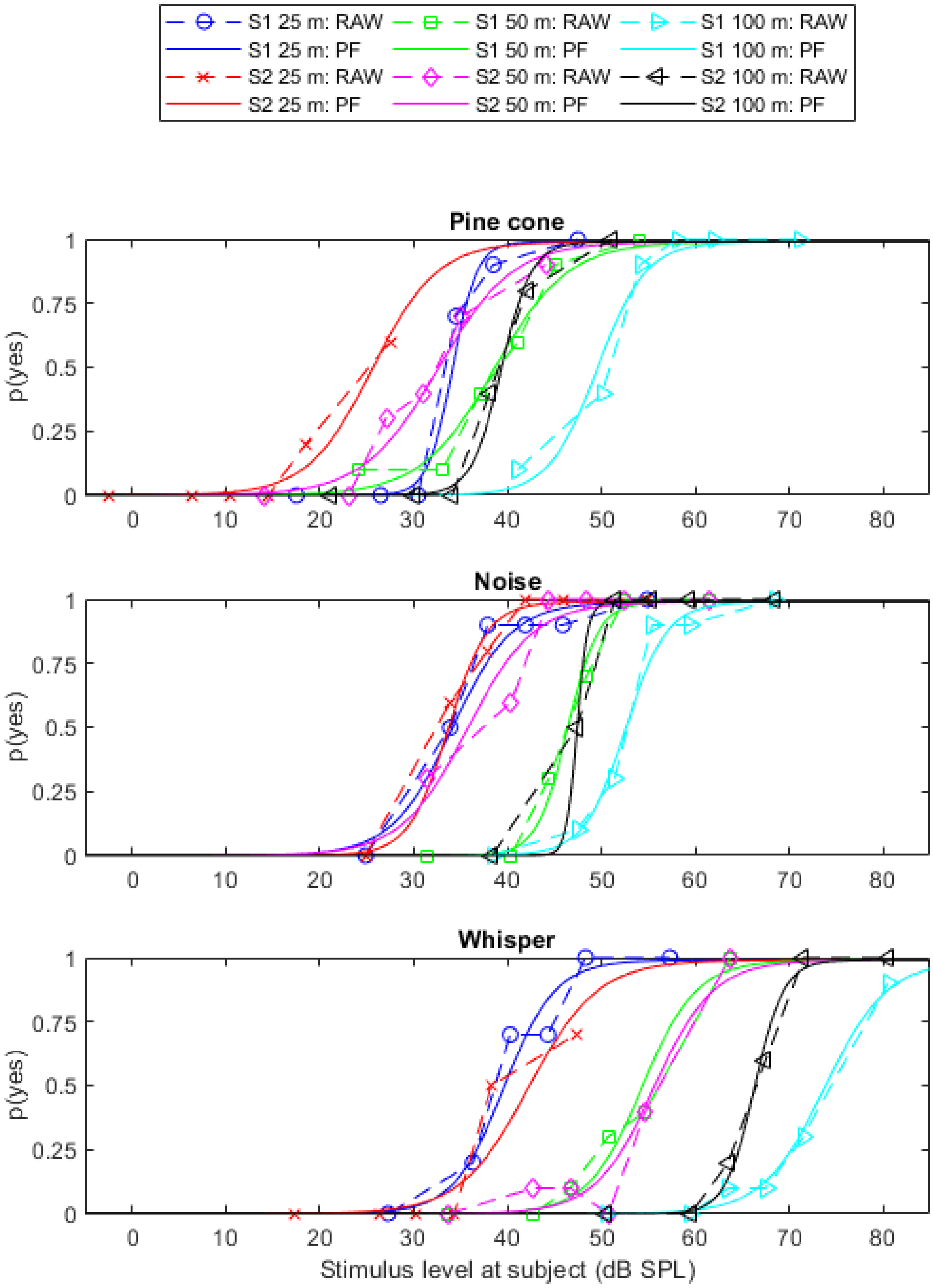

Figure 8.9 - Experiment 2. Individual results and psychometric functions for participant number 3. See Appendix E overview for full details and legend.

Table 8-8 - Experiment 2. Individual psychometric function parameter estimates for participant number 3. See Appendix E overview for full details.

Stimulus	Target distance (m)	Session	Location (dB SPL)	Location SE (dB)	Slope (β)	Slope SE (β)	Deviation	pDev
PineCone	25	1	34.12	0.66	0.76	1.03	2.86	0.11
		2	25.62	2.13	0.29	0.96	1.80	0.37
	50	1	38.61	1.43	0.25	0.09	2.90	0.57
		2	32.60	1.58	0.25	0.21	2.97	0.60
	100	1	49.55	1.31	0.39	0.40	2.80	0.31
		2	39.36	0.83	0.62	0.77	1.31	0.27
Noise	25	1	34.08	1.41	0.31	0.18	3.75	0.36
		2	33.68	1.04	0.48	0.50	1.53	0.37
	50	1	46.40	0.84	0.62	0.55	1.57	0.39
		2	35.73	2.09	0.28	0.30	4.51	0.19
	100	1	52.56	0.94	0.45	0.31	1.82	0.49
		2	47.40	0.43	1.75	0.11	0.02	0.40
Whisper	25	1	39.66	1.10	0.37	0.17	3.30	0.30
		2	42.29	1.99	0.27	0.55	6.51	0.06
	50	1	54.34	1.29	0.34	0.61	2.32	0.46
		2	55.64	1.50	0.32	0.70	7.68	0.05
	100	1	73.82	1.73	0.30	0.95	1.04	0.77
		2	66.32	0.81	0.64	0.70	1.51	0.30

8.2.1.4 *Participant 4*

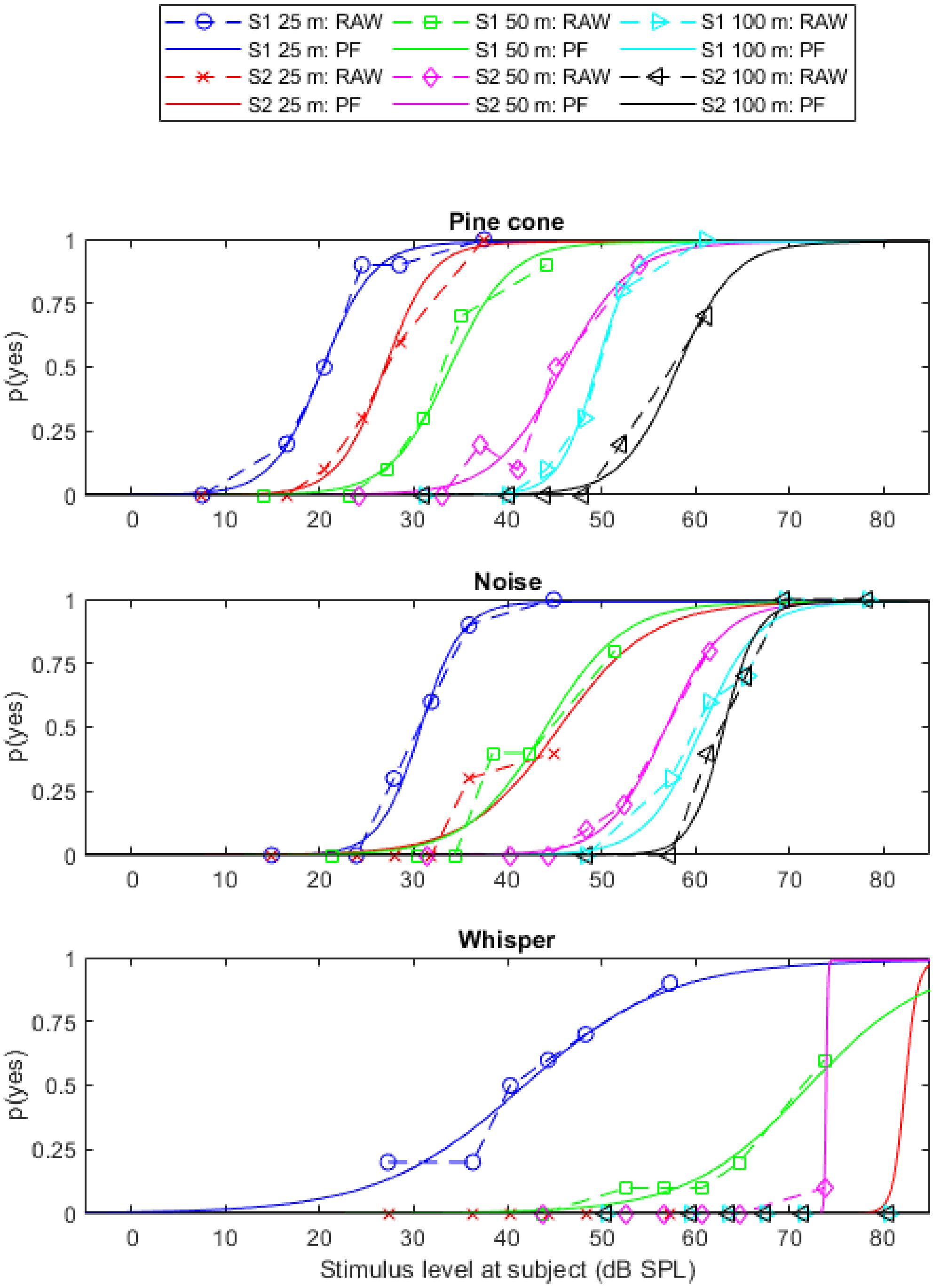

Figure 8.10 - Experiment 2. Individual results and psychometric functions for participant number 4. See Appendix E overview for full details and legend.

Table 8-9 - Experiment 2. Individual psychometric function parameter estimates for participant number 4. See Appendix E overview for full details.

Stimulus	Target distance (m)	Session	Location (dB SPL)	Location SE (dB)	Slope (β)	Slope SE (β)	Deviation	pDev
PineCone	25	1	20.31	1.12	0.37	0.16	1.17	0.75
		2	26.93	1.03	0.40	0.49	0.87	0.73
	50	1	33.86	1.30	0.31	0.22	1.89	0.65
		2	45.79	1.74	0.27	0.38	2.99	0.47
	100	1	49.44	0.96	0.50	0.51	0.50	0.74
		2	58.09	2.14	0.33	1.41	1.52	0.33
Noise	25	1	30.74	0.95	0.45	0.43	1.36	0.62
		2	45.44	11.20	0.20	0.80	4.40	0.20
	50	1	44.00	1.98	0.24	0.34	4.89	0.25
		2	56.81	2.21	0.32	1.23	0.69	0.76
	100	1	60.68	1.17	0.33	0.15	2.63	0.46
		2	62.99	0.89	0.56	0.40	2.44	0.22
Whisper	25	1	41.62	2.67	0.13	0.05	1.31	0.90
		2	82.30	0.00	1.50	0.00	0.00	0.00
	50	1	72.00	3.73	0.16	0.06	1.09	0.92
		2	73.90	1.64	11.07	0.43	0.00	0.00
	100	1	95.50	0.00	2.50	0.00	0.00	0.00
		2	95.50	0.00	2.50	0.00	0.00	0.00

8.2.1.5 *Participant 5*

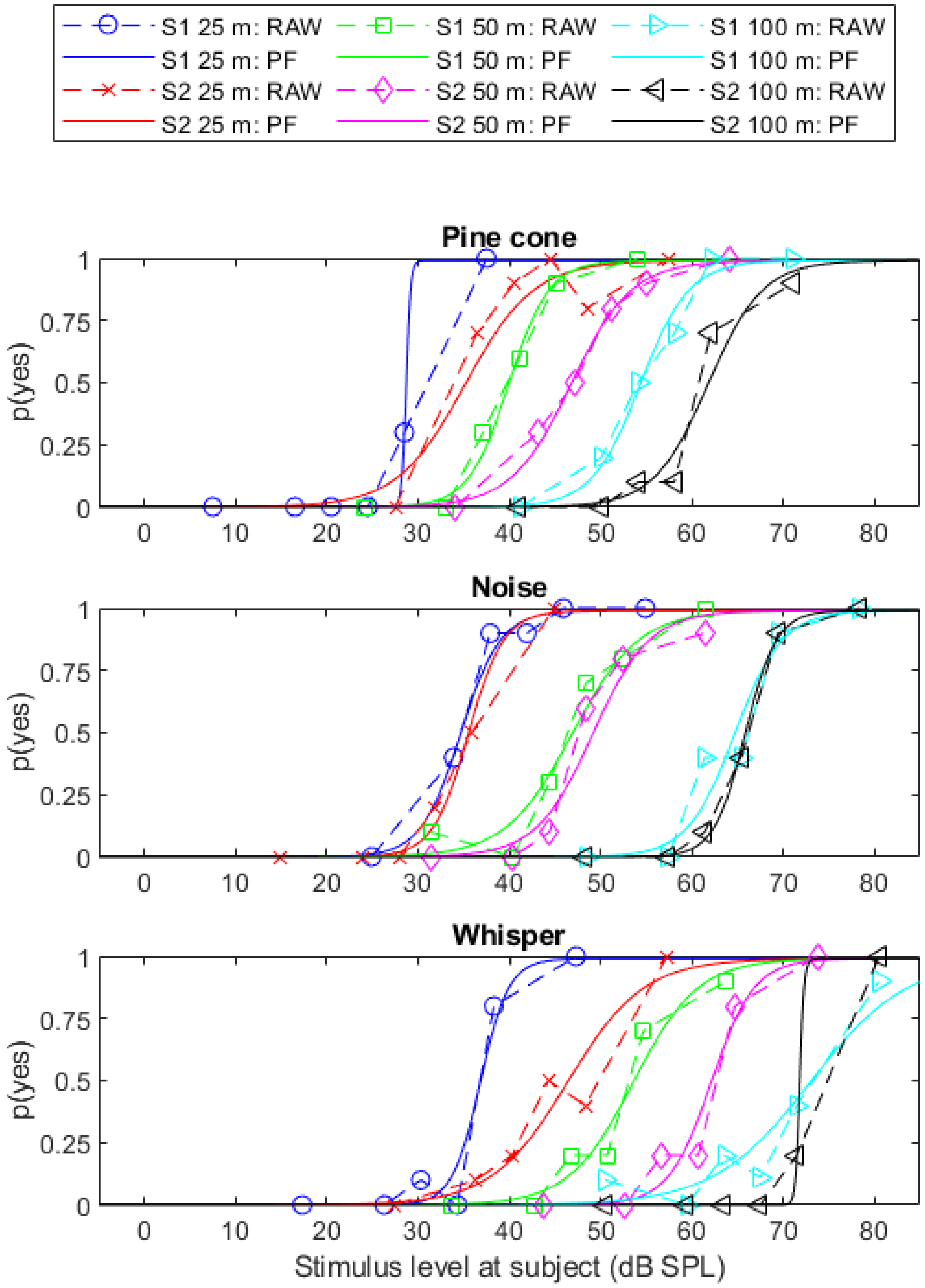

Figure 8.11 - Experiment 2. Individual results and psychometric functions for participant number 5. See Appendix E overview for full details and legend.

Table 8-10 - Experiment 2. Individual psychometric function parameter estimates for participant number 5. See Appendix E overview for full details.

Stimulus	Distance (m)	Session	Location (dB SPL)	LocationSE (dB)	Slope (β)	SlopeSE (β)	Deviation	pDev
PineCone	25	1	28.70	0.92	4.18	0.85	0.00	0.37
		2	35.18	1.73	0.27	0.14	10.06	0.01
	50	1	39.94	1.02	0.45	0.42	1.36	0.64
		2	46.78	1.26	0.30	0.10	0.79	0.92
	100	1	54.36	1.01	0.38	0.27	1.93	0.55
		2	61.79	1.33	0.34	0.69	3.46	0.30
Noise	25	1	34.65	1.06	0.46	0.35	1.75	0.37
		2	35.51	0.95	0.51	1.37	1.03	0.35
	50	1	46.88	1.33	0.27	0.10	6.26	0.14
		2	49.14	1.20	0.32	0.18	4.71	0.21
	100	1	64.70	1.07	0.38	0.18	4.23	0.24
		2	65.85	0.89	0.58	0.81	0.35	0.70
Whisper	25	1	36.83	0.96	0.64	1.23	6.17	0.01
		2	46.40	1.62	0.24	0.09	4.13	0.33
	50	1	53.60	1.42	0.28	0.34	2.96	0.51
		2	62.14	1.07	0.43	0.42	2.90	0.30
	100	1	72.99	2.17	0.19	0.07	5.99	0.17
		2	71.83	1.30	4.24	1.25	0.00	0.31

8.2.1.6 *Participant 6*

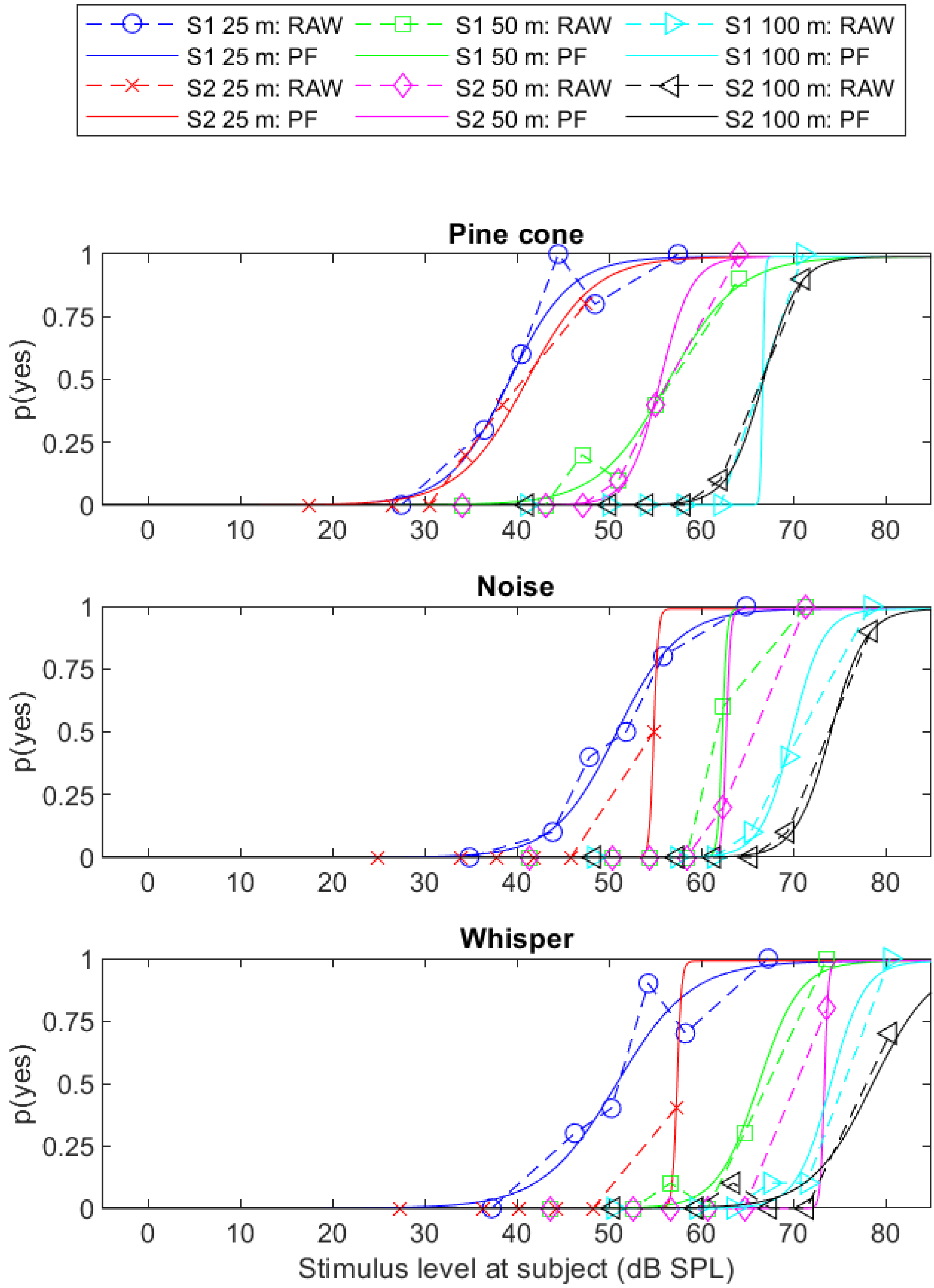

Figure 8.12 - Experiment 2. Individual results and psychometric functions for participant number 6. See Appendix E overview for full details and legend.

Table 8-11 - Experiment 2. Individual psychometric function parameter estimates for participant number 6. See Appendix E overview for full details.

Stimulus	Distance (m)	Session	Location (dB SPL)	LocationSE (dB)	Slope (ß)	SlopeSE (ß)	Deviation	pDev
PineCone	25	1	39.21	1.24	0.30	0.10	6.73	0.10
		2	41.08	1.79	0.27	0.67	2.14	0.55
	50	1	56.47	1.65	0.26	0.39	2.85	0.48
		2	55.58	0.84	0.58	1.77	0.47	0.32
	100	1	66.70	0.00	8.14	0.00	0.00	0.00
		2	66.75	2.68	0.52	3.40	0.30	0.33
Noise	25	1	50.83	1.22	0.29	0.11	1.27	0.80
		2	54.90	0.56	4.21	0.10	0.00	0.52
	50	1	62.30	0.33	4.04	0.13	0.00	0.39
		2	62.73	1.27	4.24	1.21	0.00	0.32
	100	1	69.88	0.85	0.58	1.78	0.47	0.35
		2	74.05	2.72	0.52	2.97	0.30	0.28
Whisper	25	1	50.81	1.47	0.25	0.13	5.44	0.23
		2	57.40	0.91	4.09	0.14	0.00	0.00
	50	1	66.23	1.26	0.44	1.69	4.72	0.06
		2	73.40	1.21	4.62	1.02	0.00	0.20
	100	1	74.06	1.88	0.55	2.35	2.58	0.11
		2	78.55	1.90	0.30	0.98	5.96	0.05

8.2.1.7 *Participant 7*

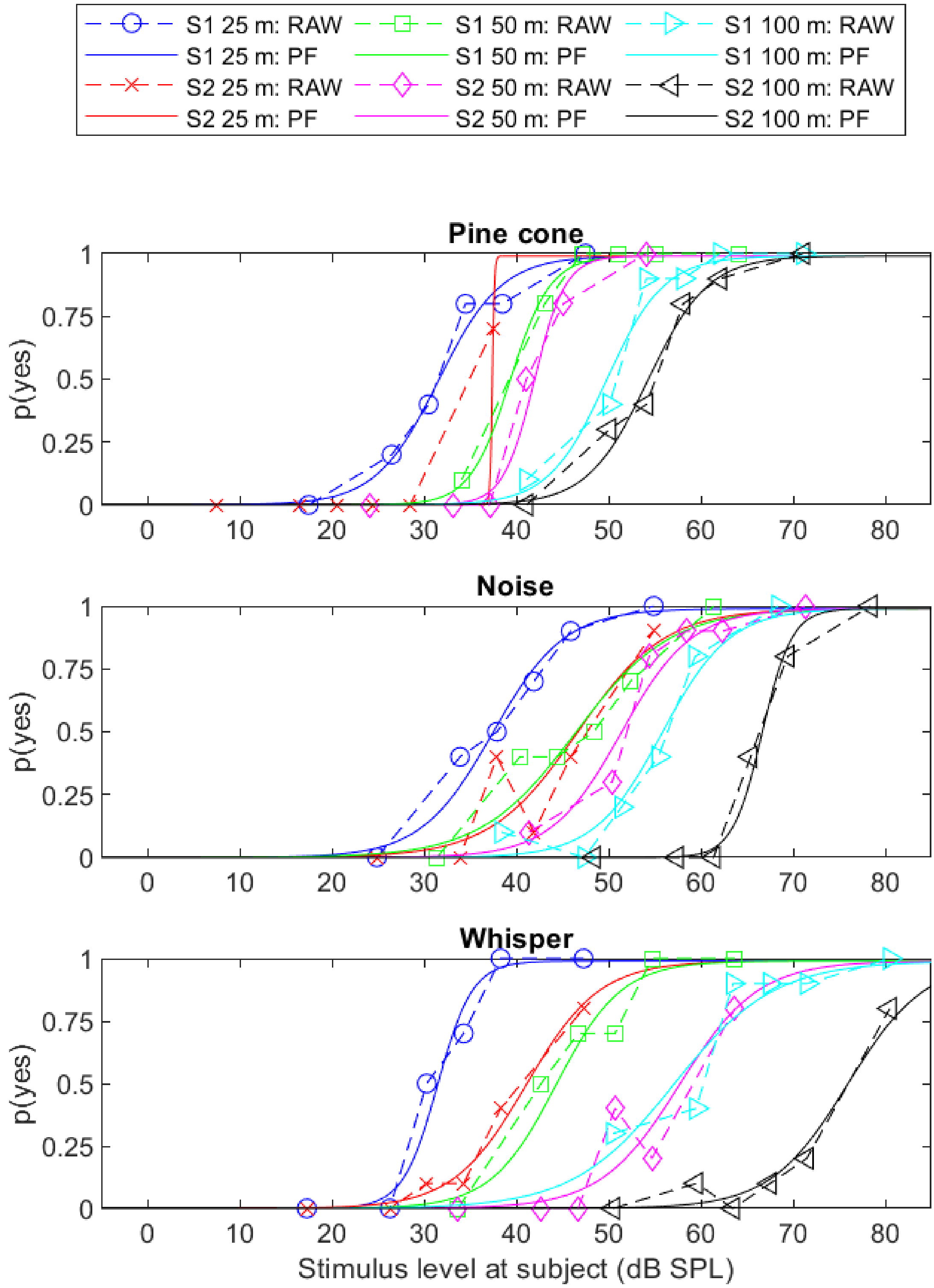

Figure 8.13 - Experiment 2. Individual results and psychometric functions for participant number 7. See Appendix E overview for full details and legend.

Table 8-12 - Experiment 2. Individual psychometric function parameter estimates for participant number 7. See Appendix E overview for full details.

Stimulus	**Distance (m)**	**Session**	**Location (dB SPL)**	**LocationSE (dB)**	**Slope (ß)**	**SlopeSE (ß)**	**Deviation**	**pDev**
PineCone	25	1	31.57	1.23	0.29	0.10	1.62	0.75
		2	37.40	0.54	8.56	0.86	0.00	0.63
	50	1	39.33	1.99	0.46	0.54	0.87	0.32
		2	41.95	0.89	0.59	0.76	2.18	0.21
	100	1	49.77	1.53	0.32	0.26	2.21	0.54
		2	54.26	1.21	0.31	0.14	1.25	0.78
Noise	25	1	37.45	1.50	0.26	0.10	1.81	0.73
		2	46.59	1.93	0.21	0.29	7.87	0.09
	50	1	46.40	1.56	0.19	0.07	3.98	0.47
		2	51.36	1.46	0.26	0.10	2.15	0.67
	100	1	55.80	1.42	0.27	0.16	6.28	0.17
		2	66.66	0.93	0.62	0.84	1.31	0.29
Whisper	25	1	31.48	0.92	0.53	0.44	3.55	0.16
		2	41.08	1.95	0.25	0.57	1.24	0.77
	50	1	44.33	1.32	0.28	0.11	4.33	0.31
		2	57.69	1.99	0.23	0.58	6.18	0.11
	100	1	57.06	2.62	0.18	0.06	3.77	0.47
		2	75.89	2.24	0.24	0.40	3.15	0.39

8.2.1.8 *Participant 8*

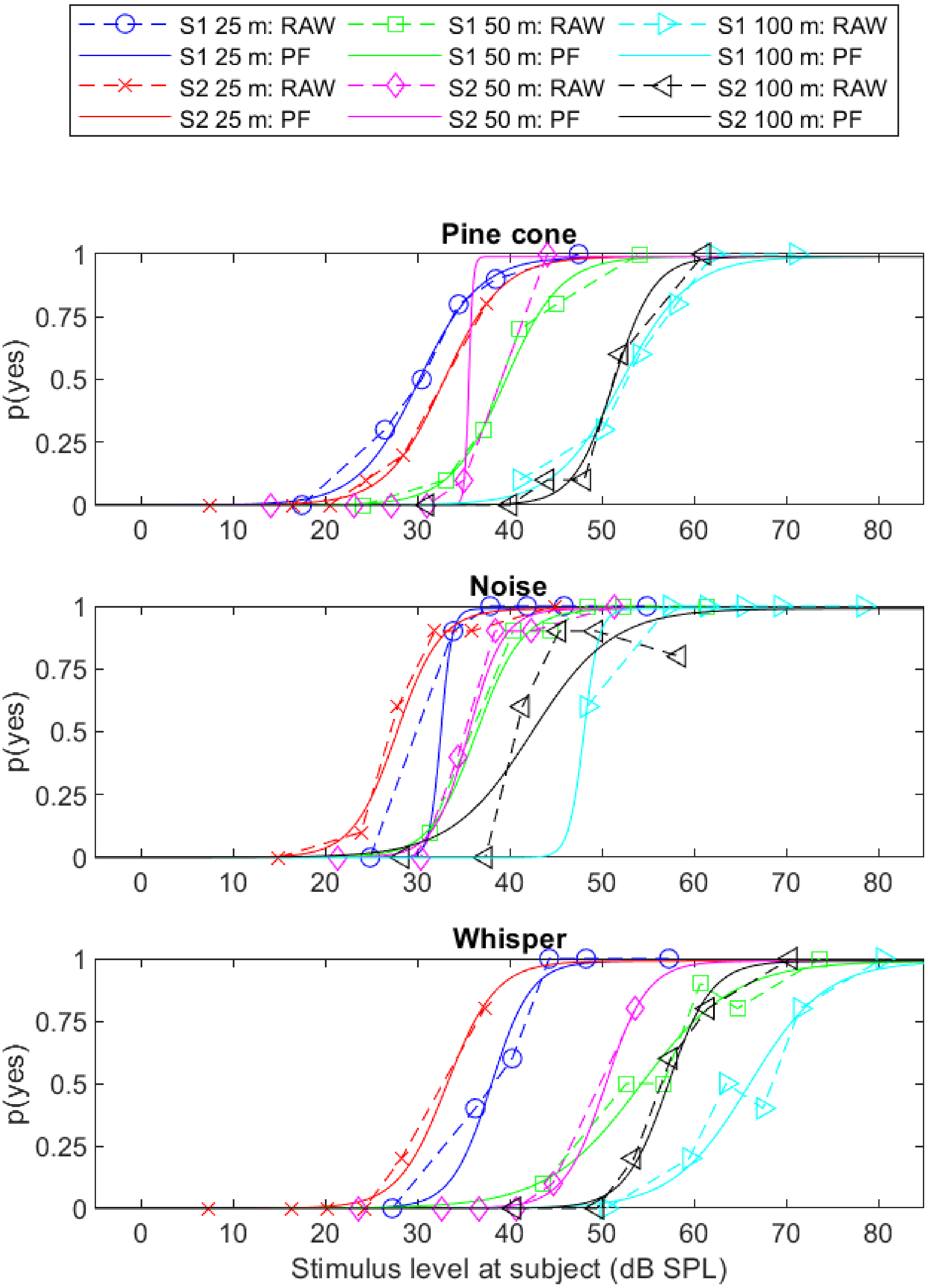

Figure 8.14 - Experiment 2. Individual results and psychometric functions for participant number 8. See Appendix E overview for full details and legend.

Table 8-13 - Experiment 2. Individual psychometric function parameter estimates for participant number 8. See Appendix E overview for full details.

Stimulus	Distance (m)	Session	Location (dB SPL)	LocationSE (dB)	Slope (ß)	SlopeSE (ß)	Deviation	pDev
PineCone	25	1	30.18	1.30	0.30	0.19	0.79	0.91
		2	32.91	2.12	0.32	1.01	0.69	0.78
	50	1	39.55	1.21	0.33	0.23	0.83	0.89
		2	35.61	2.28	4.33	2.47	0.00	0.26
	100	1	52.19	1.47	0.27	0.13	2.18	0.69
		2	51.35	1.01	0.46	0.97	1.75	0.40
Noise	25	1	32.56	0.27	1.64	0.02	0.00	0.58
		2	27.77	1.04	0.42	0.29	1.94	0.48
	50	1	36.42	1.88	0.40	0.49	1.31	0.42
		2	35.58	0.95	0.53	0.37	3.17	0.22
	100	1	48.06	25.06	1.21	0.11	0.00	0.38
		2	42.38	1.46	0.23	0.08	15.86	0.00
Whisper	25	1	38.02	1.01	0.47	0.34	2.69	0.25
		2	33.23	2.16	0.38	1.66	1.22	0.33
	50	1	54.25	1.85	0.20	0.07	2.87	0.65
		2	50.42	2.19	0.44	2.26	0.42	0.37
	100	1	66.00	1.44	0.24	0.08	3.47	0.48
		2	57.20	1.05	0.42	0.39	1.36	0.61

8.2.1.9 *Participant 9*

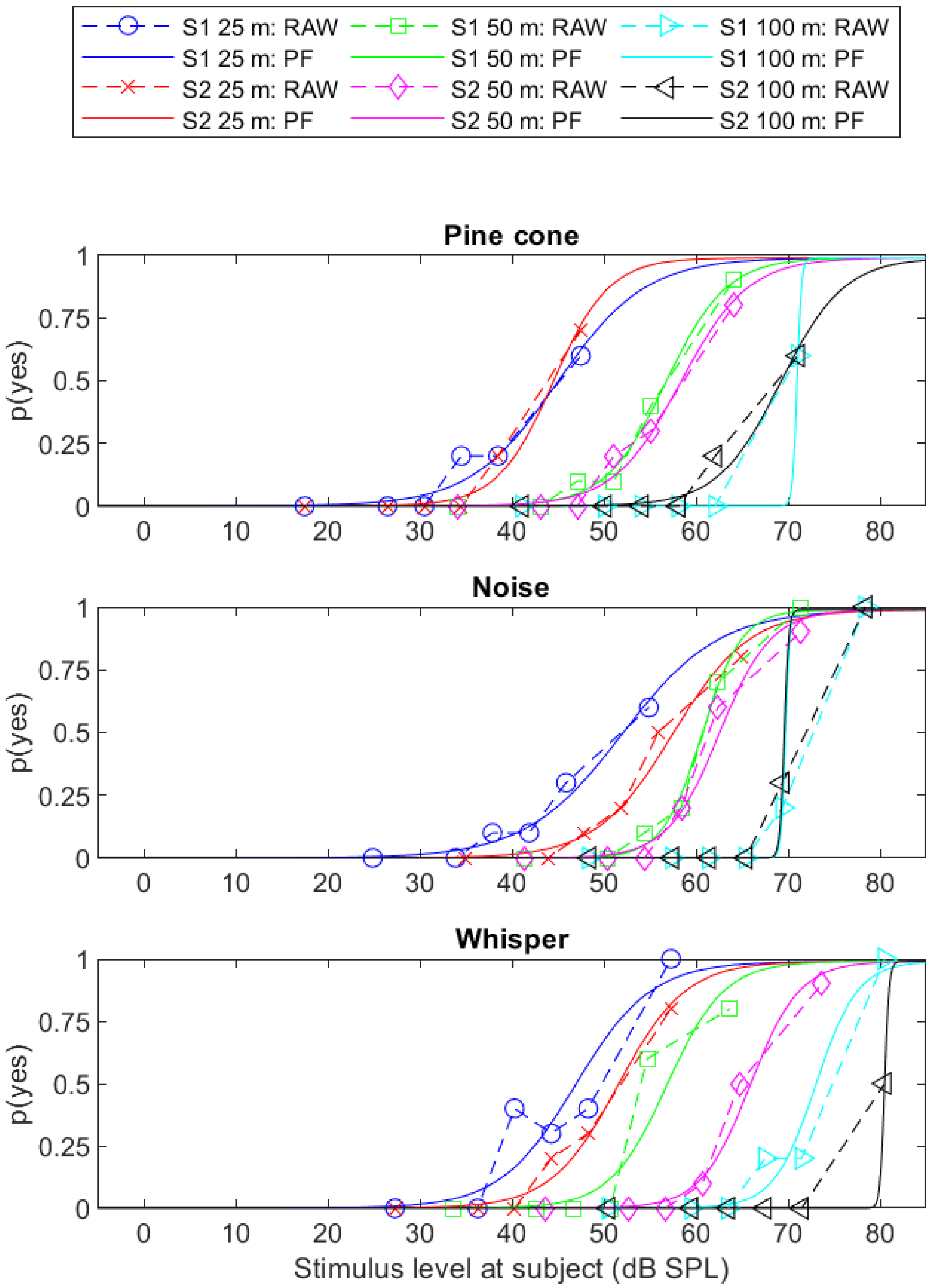

Figure 8.15 - Experiment 2. Individual results and psychometric functions for participant number 9. See Appendix E overview for full details and legend.

Table 8-14 - Experiment 2. Individual psychometric function parameter estimates for participant number 9. See Appendix E overview for full details.

Stimulus	Distance (m)	Session	Location (dB SPL)	LocationSE (dB)	Slope (ß)	SlopeSE (ß)	Deviation	pDev
PineCone	25	1	44.99	3.03	0.21	0.54	2.34	0.61
		2	44.49	2.12	0.33	1.37	1.52	0.33
	50	1	56.75	1.65	0.30	0.80	0.99	0.81
		2	58.36	1.83	0.27	0.39	1.79	0.61
	100	1	71.00	0.39	4.14	0.39	0.00	0.00
		2	69.22	2.01	0.29	0.99	1.80	0.36
Noise	25	1	52.01	2.79	0.19	0.32	1.25	0.82
		2	57.50	1.82	0.24	0.22	1.44	0.82
	50	1	60.68	1.07	0.47	0.68	0.88	0.61
		2	62.53	1.29	0.36	0.64	2.51	0.45
	100	1	69.73	1.35	4.24	1.28	0.00	0.34
		2	69.60	0.74	4.18	0.65	0.00	0.30
Whisper	25	1	46.92	1.62	0.24	0.15	7.68	0.08
		2	51.56	1.95	0.27	0.81	1.79	0.64
	50	1	56.87	1.60	0.32	0.98	7.42	0.04
		2	65.91	1.44	0.37	1.24	1.73	0.51
	100	1	73.03	1.40	0.44	1.70	3.42	0.20
		2	80.50	0.55	4.21	0.22	0.00	0.54

8.2.1.10 *Participant 10*

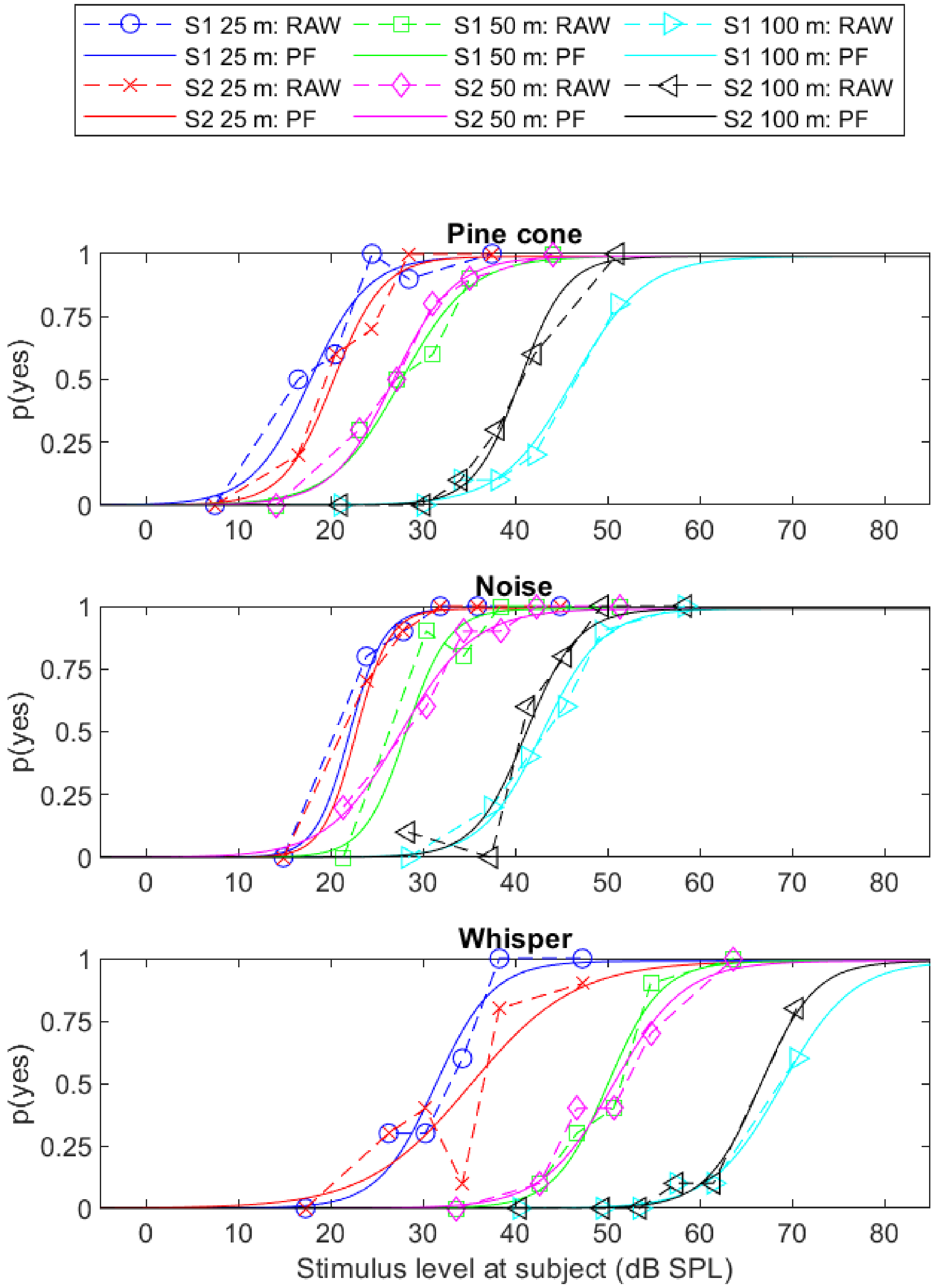

Figure 8.16 - Experiment 2. Individual results and psychometric functions for participant number 10. See Appendix E overview for full details and legend.

Table 8-15 - Experiment 2. Individual psychometric function parameter estimates for participant number 10. See Appendix E overview for full details.

Stimulus	Distance (m)	Session	Location (dB SPL)	LocationSE (dB)	Slope (ß)	SlopeSE (ß)	Deviation	pDev
PineCone	25	1	17.78	1.34	0.32	0.15	4.88	0.16
		2	20.31	1.12	0.37	0.28	2.36	0.48
	50	1	27.65	1.36	0.27	0.10	1.75	0.75
		2	26.78	1.26	0.30	0.15	0.79	0.90
	100	1	46.26	2.05	0.26	0.60	1.17	0.72
		2	40.53	1.13	0.40	0.46	0.87	0.77
Noise	25	1	22.08	1.35	0.54	0.60	1.43	0.23
		2	22.77	1.02	0.55	0.61	0.82	0.31
	50	1	28.18	1.64	0.44	0.57	5.29	0.06
		2	27.62	1.92	0.25	0.22	1.24	0.79
	100	1	42.94	1.27	0.30	0.10	0.89	0.87
		2	41.19	1.09	0.36	0.18	8.80	0.05
Whisper	25	1	31.40	1.23	0.32	0.11	4.75	0.25
		2	35.19	1.94	0.18	0.06	9.77	0.07
	50	1	50.03	1.15	0.34	0.17	1.88	0.58
		2	50.65	1.37	0.26	0.16	2.31	0.62
	100	1	68.92	2.48	0.26	0.92	1.01	0.66
		2	66.59	2.14	0.34	1.42	1.09	0.49

8.2.1.11 *Participant 11*

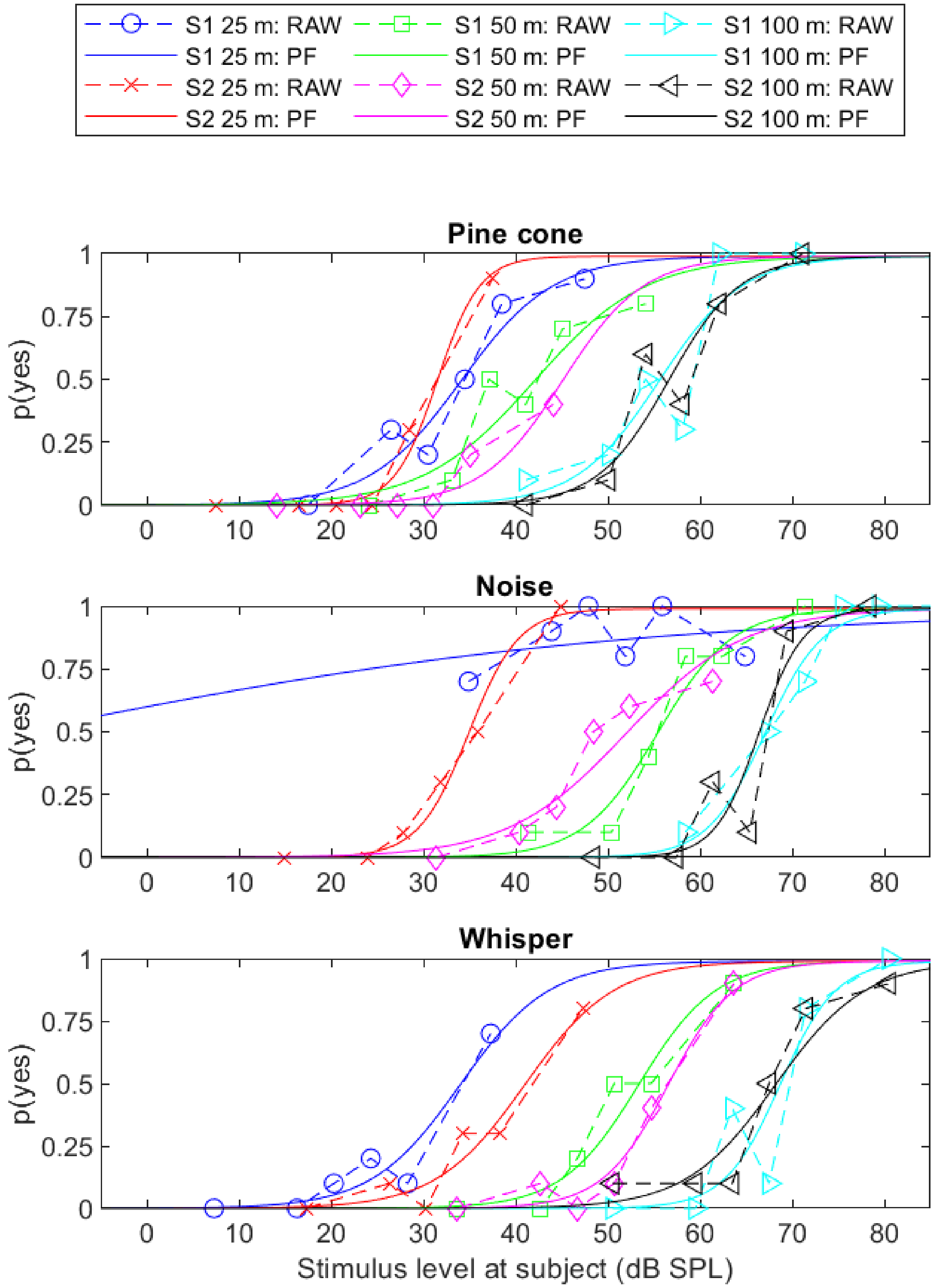

Figure 8.17 - Experiment 2. Individual results and psychometric functions for participant number 11. See Appendix E overview for full details and legend.

Table 8-16 - Experiment 2. Individual psychometric function parameter estimates for participant number 11. See Appendix E overview for full details.

Stimulus	Distance (m)	Session	Location (dB SPL)	LocationSE (dB)	Slope (ß)	SlopeSE (ß)	Deviation	pDev
PineCone	25	1	34.12	1.65	0.20	0.07	3.02	0.62
		2	31.51	1.93	0.43	1.97	1.67	0.28
	50	1	42.08	2.09	0.17	0.05	4.12	0.48
		2	45.10	4.37	0.23	1.11	2.42	0.38
	100	1	55.58	1.54	0.22	0.08	10.54	0.03
		2	56.59	1.36	0.26	0.13	5.38	0.22
Noise	25	1	-14.55	2934.23	0.03	0.07	7.91	0.19
		2	34.92	1.31	0.37	0.43	1.44	0.68
	50	1	55.41	1.39	0.25	0.15	3.48	0.44
		2	52.03	2.44	0.16	0.05	2.80	0.62
	100	1	67.07	1.54	0.32	0.26	2.60	0.50
		2	66.41	1.03	0.42	0.40	8.74	0.01
Whisper	25	1	34.00	2.80	0.21	0.37	2.71	0.53
		2	41.01	2.12	0.21	0.08	3.49	0.48
	50	1	53.35	1.68	0.24	0.09	3.06	0.54
		2	56.55	1.76	0.28	0.57	3.32	0.38
	100	1	68.69	1.21	0.33	0.31	9.92	0.02
		2	68.19	1.73	0.21	0.06	4.40	0.37

8.2.1.12 *Participant 12*

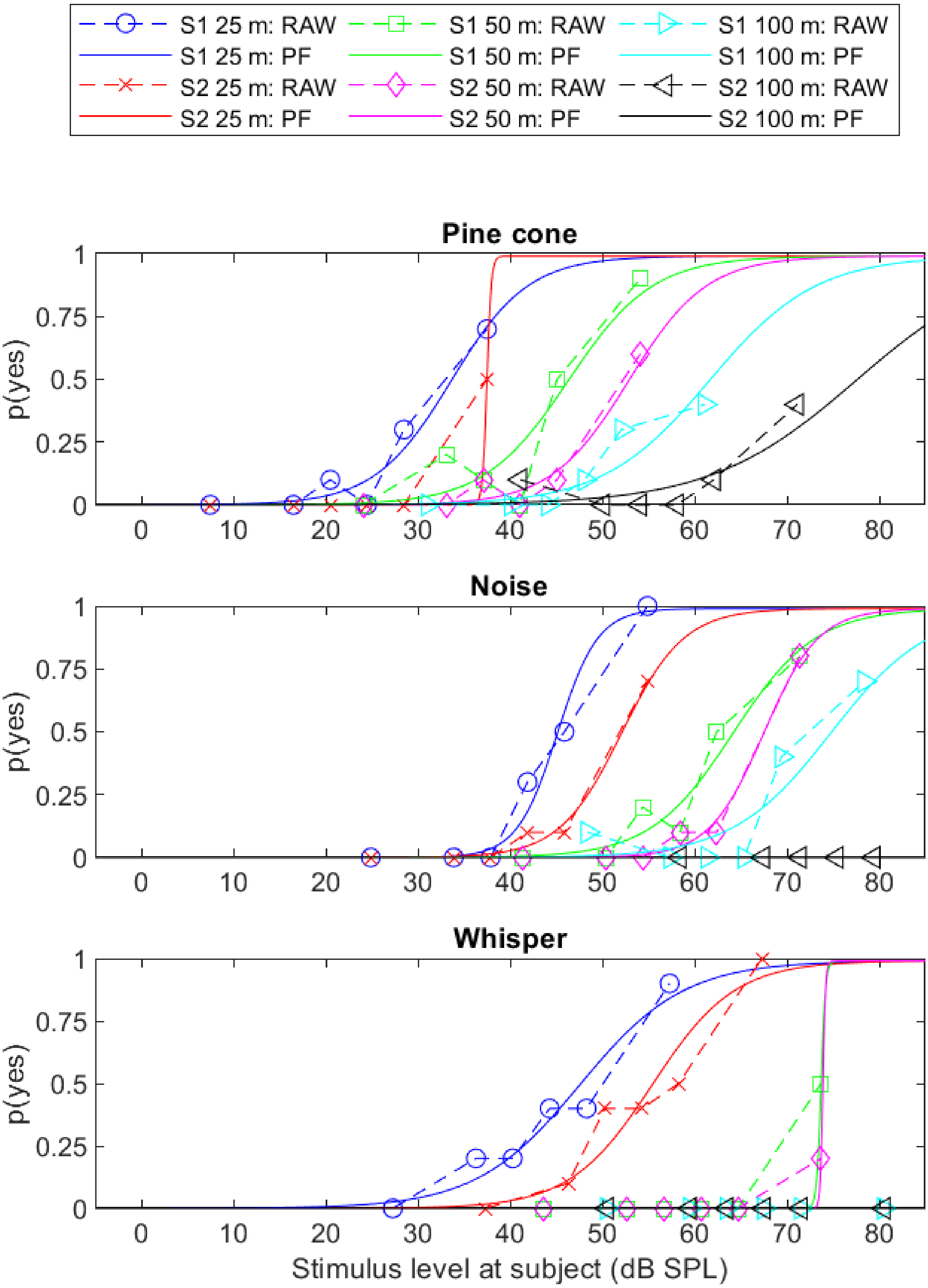

Figure 8.18 - Experiment 2. Individual results and psychometric functions for participant number 12. See Appendix E overview for full details and legend.

Table 8-17 - Experiment 2. Individual psychometric function parameter estimates for participant number 12. See Appendix E overview for full details.

Stimulus	Distance (m)	Session	Location (dB SPL)	LocationSE (dB)	Slope (ß)	SlopeSE (ß)	Deviation	pDev
PineCone	25	1	33.64	2.22	0.24	0.56	3.43	0.34
		2	37.50	0.17	4.21	0.10	0.00	0.50
	50	1	46.32	1.89	0.22	0.07	8.79	0.06
		2	52.87	2.89	0.24	0.95	3.00	0.31
	100	1	61.61	5.22	0.17	0.57	2.97	0.48
		2	77.47	502.67	0.13	0.47	6.62	0.07
Noise	25	1	45.15	1.13	0.46	0.95	2.13	0.27
		2	52.10	2.25	0.30	1.04	1.02	0.59
	50	1	64.19	2.08	0.22	0.21	3.30	0.45
		2	67.49	2.28	0.34	1.43	1.09	0.47
	100	1	74.75	3.08	0.19	0.07	9.92	0.03
		2	103.40	0.00	2.50	0.00	0.00	0.00
Whisper	25	1	47.58	2.00	0.18	0.06	2.09	0.75
		2	55.43	1.55	0.22	0.07	4.03	0.38
	50	1	73.70	0.27	4.21	0.22	0.00	0.54
		2	73.90	1.81	6.95	0.37	0.00	0.00
	100	1	95.50	0.00	2.50	0.00	0.00	0.00
		2	95.50	0.00	2.50	0.00	0.00	0.00

8.2.1.13 *Participant 13*

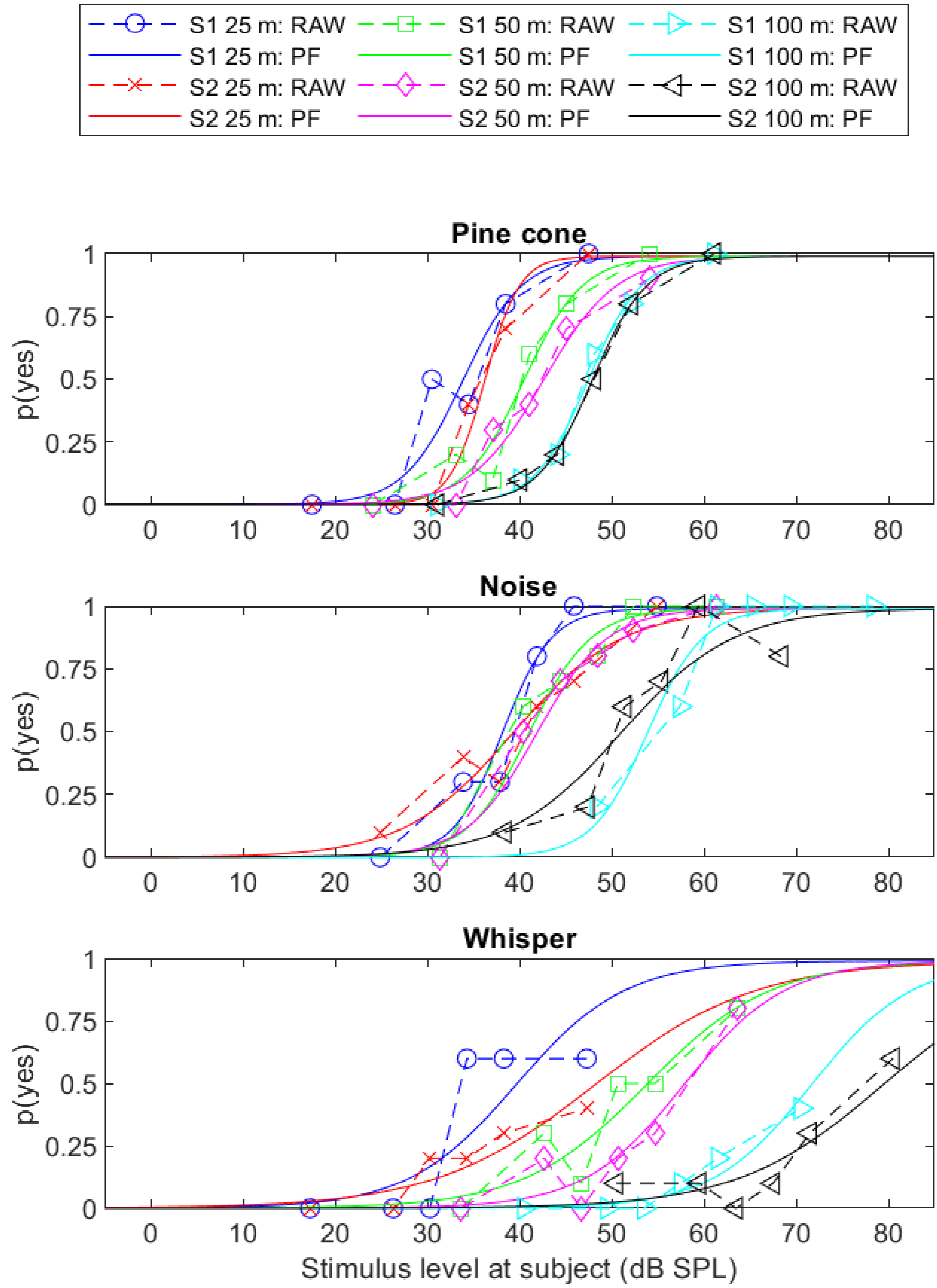

Figure 8.19 - Experiment 2. Individual results and psychometric functions for participant number 13. See Appendix E overview for full details and legend.

Table 8-18 - Experiment 2. Individual psychometric function parameter estimates for participant number 13. See Appendix E overview for full details.

Stimulus	**Distance (m)**	**Session**	**Location (dB SPL)**	**LocationSE (dB)**	**Slope (ß)**	**SlopeSE (ß)**	**Deviation**	**pDev**
PineCone	25	1	33.87	1.21	0.31	0.20	5.86	0.17
		2	36.25	0.97	0.53	0.60	1.96	0.33
	50	1	40.47	1.25	0.31	0.11	3.21	0.40
		2	42.60	1.50	0.25	0.09	2.97	0.54
	100	1	47.43	1.19	0.34	0.17	0.59	0.95
		2	47.89	1.17	0.34	0.12	0.48	0.95
Noise	25	1	38.16	1.05	0.38	0.26	3.58	0.29
		2	39.31	2.01	0.17	0.06	3.01	0.66
	50	1	41.05	1.44	0.30	0.15	3.62	0.34
		2	41.90	1.57	0.26	0.09	2.21	0.64
	100	1	53.88	1.87	0.33	0.37	3.71	0.19
		2	51.02	1.95	0.16	0.05	8.53	0.09
Whisper	25	1	39.65	2.50	0.17	0.06	13.10	0.01
		2	48.22	14.45	0.11	0.05	3.21	0.60
	50	1	53.79	2.47	0.15	0.06	4.34	0.43
		2	57.86	2.46	0.18	0.06	5.32	0.27
	100	1	71.69	6.92	0.18	0.74	1.66	0.65
		2	79.46	5.90	0.13	0.12	4.65	0.39

8.2.1.14 *Participant 14*

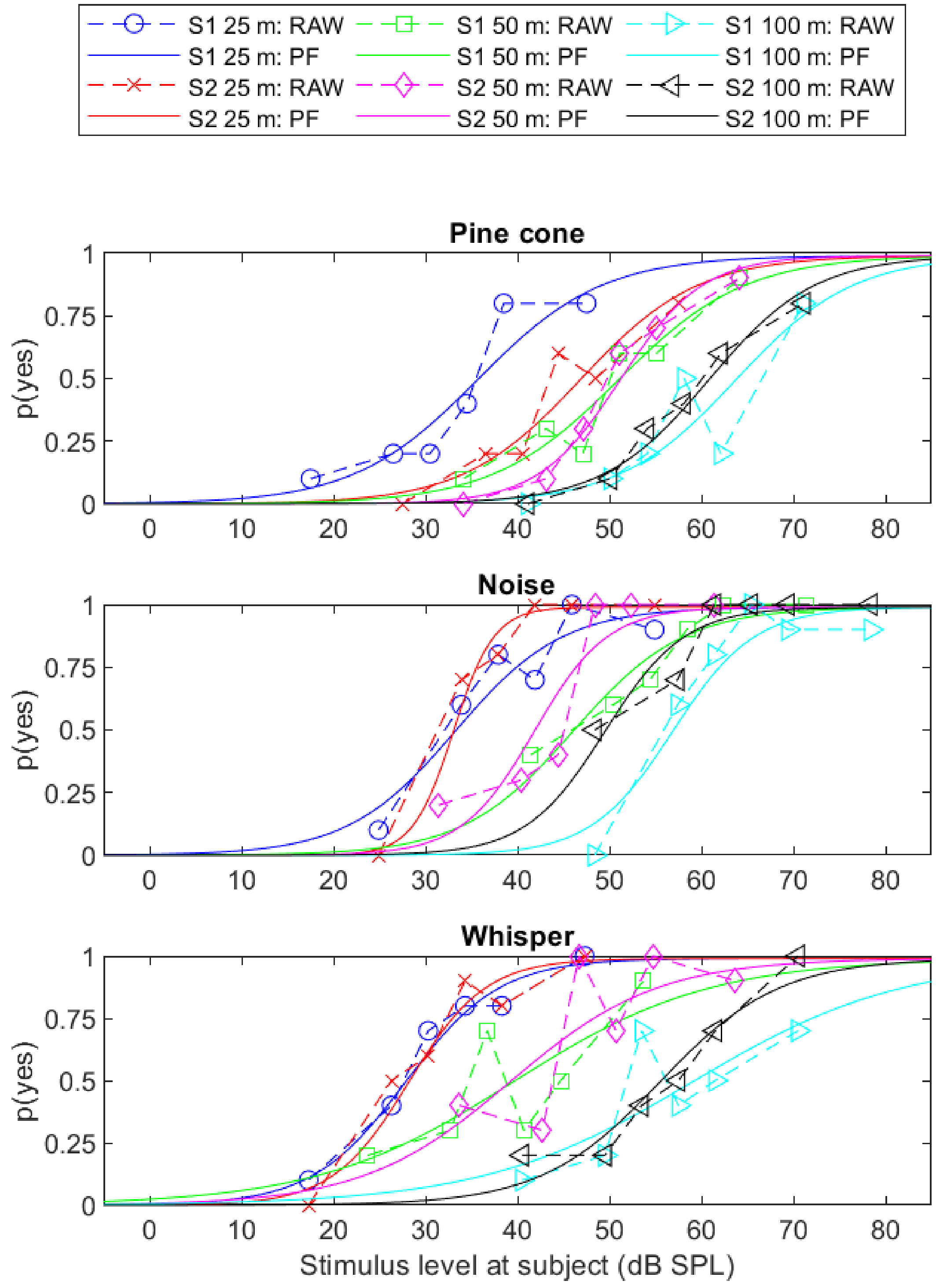

Figure 8.20 - Experiment 2. Individual results and psychometric functions for participant number 14. See Appendix E overview for full details and legend.

Table 8-19 - Experiment 2. Individual psychometric function parameter estimates for participant number 14. See Appendix E overview for full details.

Stimulus	Distance (m)	Session	Location (dB SPL)	LocationSE (dB)	Slope (ß)	SlopeSE (ß)	Deviation	pDev
PineCone	25	1	35.66	2.42	0.15	0.05	2.99	0.63
		2	46.83	2.46	0.16	0.06	3.05	0.64
	50	1	51.00	2.22	0.15	0.05	2.03	0.77
		2	51.17	1.53	0.22	0.07	1.40	0.88
	100	1	63.75	2.60	0.16	0.06	5.06	0.34
		2	60.81	2.10	0.17	0.06	1.28	0.93
Noise	25	1	33.24	2.19	0.17	0.05	5.82	0.23
		2	33.04	1.13	0.45	0.55	2.27	0.28
	50	1	46.04	3.07	0.17	0.06	2.71	0.66
		2	41.84	1.61	0.25	0.08	10.14	0.02
	100	1	56.88	1.76	0.22	0.08	9.58	0.05
		2	49.53	3.07	0.23	0.30	3.85	0.20
Whisper	25	1	27.85	1.75	0.20	0.07	1.41	0.87
		2	28.20	1.58	0.24	0.08	4.13	0.42
	50	1	39.54	3.64	0.09	0.04	6.54	0.22
		2	39.20	4.32	0.13	0.05	14.57	0.01
	100	1	59.03	4.09	0.09	0.03	5.40	0.32
		2	55.50	1.94	0.15	0.05	3.66	0.56

8.2.1.15 *Participant 15*

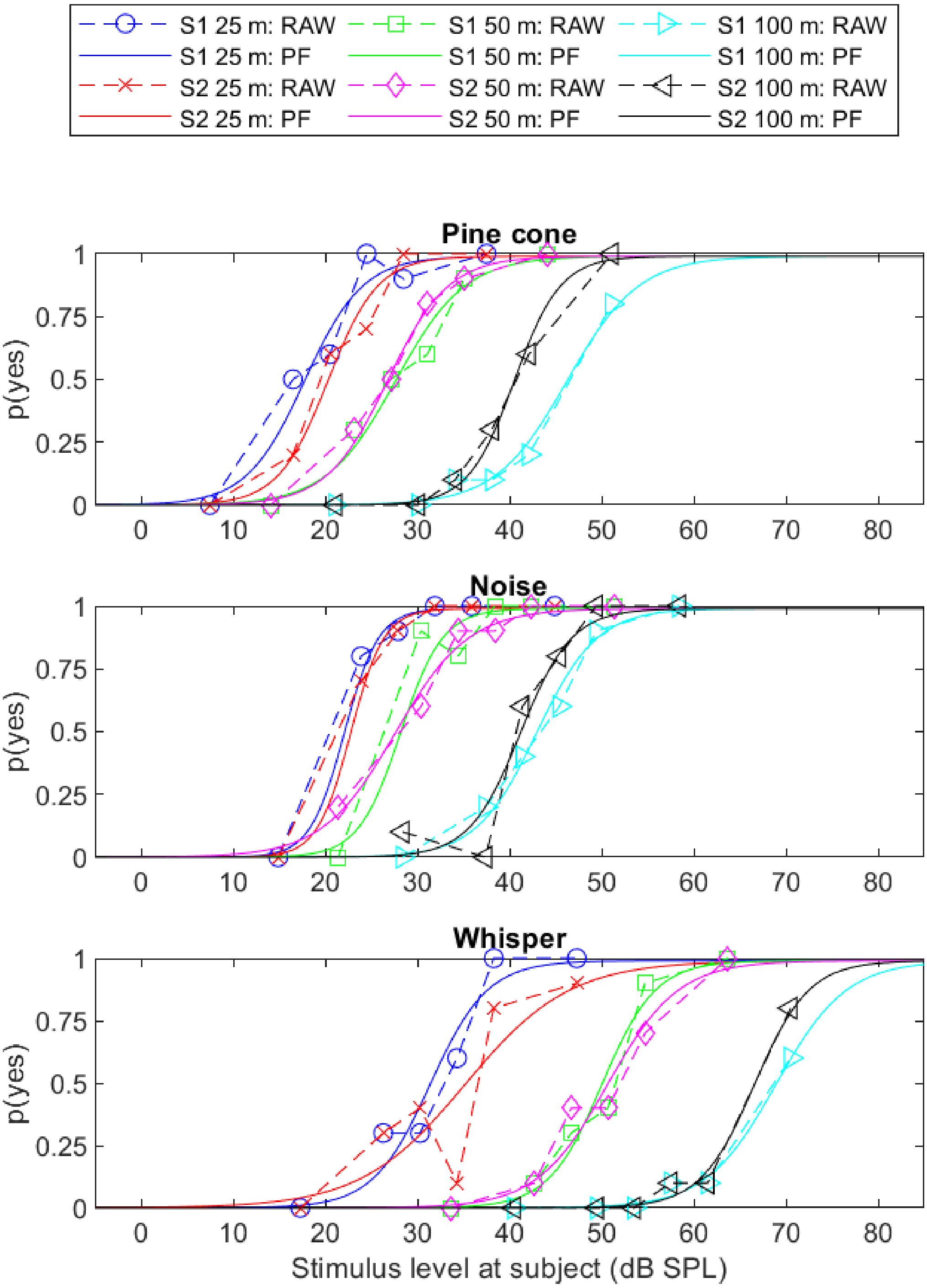

Figure 8.21 - Experiment 2. Individual results and psychometric functions for participant number 15. See Appendix E overview for full details and legend.

Table 8-20- Experiment 2. Individual psychometric function parameter estimates for participant number 15. See Appendix E overview for full details.

Stimulus	Distance (m)	Session	Location (dB SPL)	LocationSE (dB)	Slope (ß)	SlopeSE (ß)	Deviation	pDev
PineCone	25	1	17.78	1.26	0.32	0.15	4.88	0.18
		2	20.31	1.16	0.37	0.25	2.36	0.46
	50	1	27.65	1.41	0.27	0.10	1.75	0.72
		2	26.78	1.32	0.30	0.14	0.79	0.89
	100	1	46.26	1.96	0.26	0.45	1.17	0.80
		2	40.53	1.16	0.40	0.40	0.87	0.76
Noise	25	1	22.08	1.30	0.54	0.61	1.43	0.18
		2	22.77	1.18	0.55	0.60	0.82	0.34
	50	1	28.18	1.53	0.44	0.56	5.29	0.07
		2	27.62	1.93	0.25	0.19	1.24	0.81
	100	1	42.94	1.22	0.30	0.14	0.89	0.88
		2	41.19	1.06	0.36	0.19	8.80	0.05
Whisper	25	1	31.40	1.22	0.32	0.14	4.75	0.23
		2	35.19	1.90	0.18	0.06	9.77	0.05
	50	1	50.03	1.17	0.34	0.21	1.88	0.61
		2	50.65	1.36	0.26	0.09	2.31	0.70
	100	1	68.92	2.52	0.26	0.94	1.01	0.65
		2	66.59	2.16	0.34	1.83	1.09	0.55

8.2.1.16 *Participant 16*

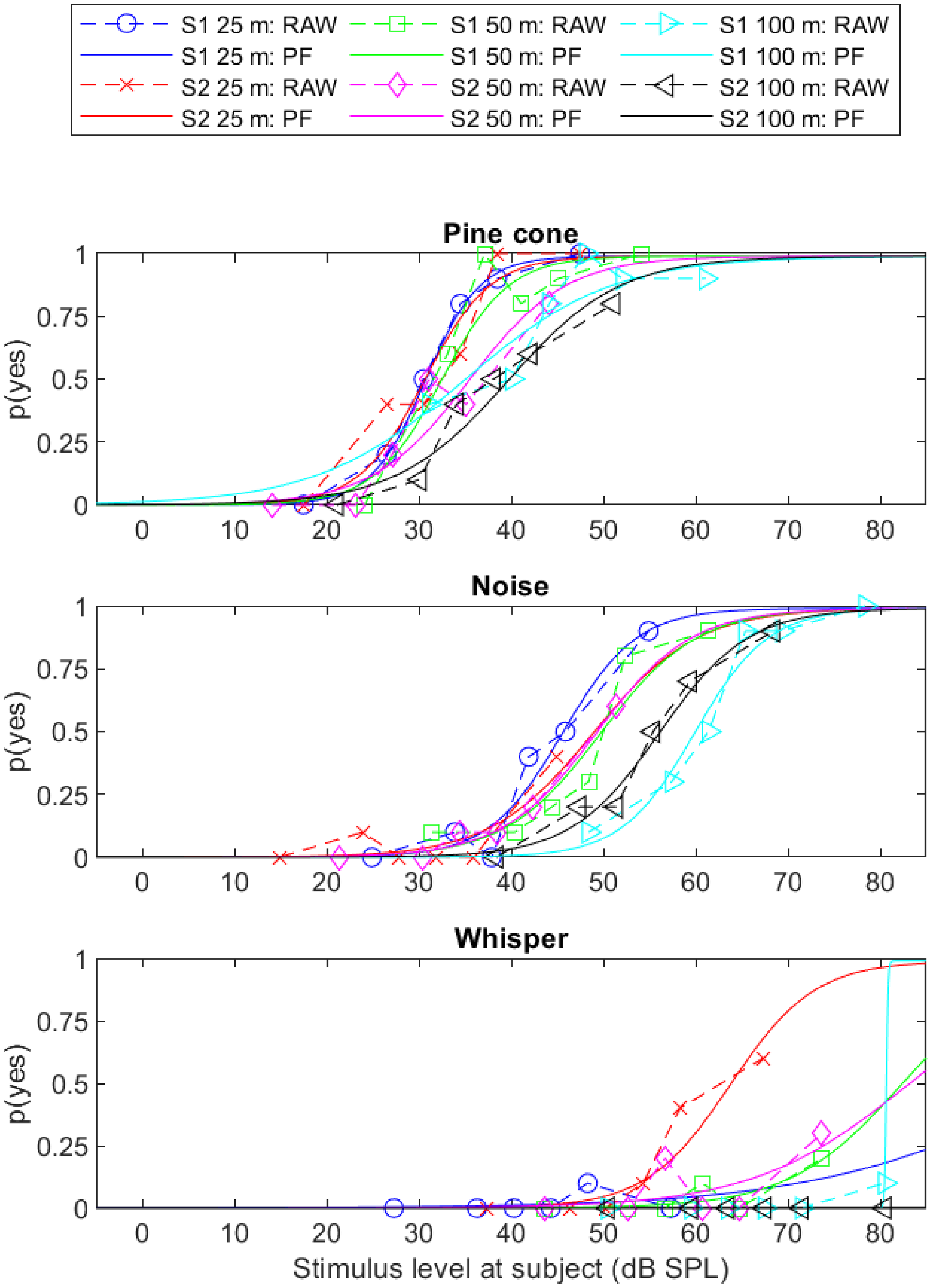

Figure 8.22 - Experiment 2. Individual results and psychometric functions for participant number 16. See Appendix E overview for full details and legend.

Table 8-21 - Experiment 2. Individual psychometric function parameter estimates for participant number 16. See Appendix E overview for full details.

Stimulus	Distance (m)	Session	Location (dB SPL)	LocationSE (dB)	Slope (ß)	SlopeSE (ß)	Deviation	pDev
PineCone	25	1	30.71	1.22	0.34	0.17	0.48	0.95
		2	30.60	1.30	0.28	0.10	5.40	0.23
	50	1	32.45	1.64	0.27	0.16	9.35	0.03
		2	35.43	2.02	0.19	0.20	4.27	0.39
	100	1	34.90	5.36	0.13	0.05	5.45	0.29
		2	39.70	2.02	0.16	0.06	2.45	0.70
Noise	25	1	45.59	1.66	0.25	0.21	3.88	0.37
		2	48.92	88.26	0.18	0.97	6.20	0.02
	50	1	49.73	1.79	0.19	0.06	3.73	0.50
		2	49.25	4.31	0.20	0.40	0.96	0.91
	100	1	59.90	1.54	0.25	0.09	1.78	0.76
		2	56.07	1.67	0.21	0.07	1.26	0.92
Whisper	25	1	100.90	2343.66	0.07	0.69	3.24	0.28
		2	63.78	2.43	0.22	0.14	3.24	0.36
	50	1	82.32	427.02	0.16	0.96	2.73	0.31
		2	83.07	726.46	0.12	0.50	7.73	0.07
	100	1	80.70	1.65	11.07	0.37	0.00	0.00
		2	95.50	0.00	2.50	0.00	0.00	0.00

8.2.1.17 *Participant 17*

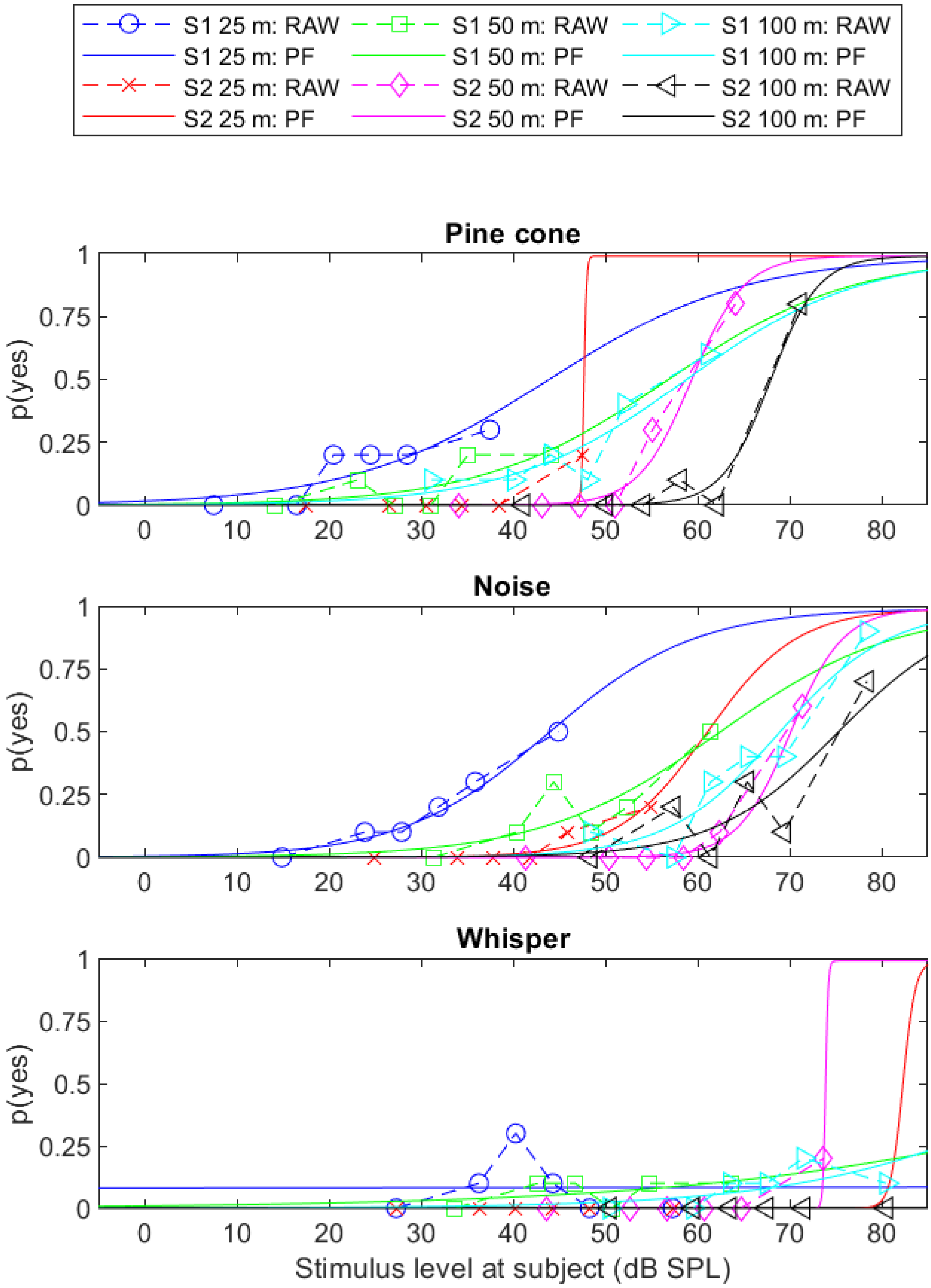

Figure 8.23 - Experiment 2. Individual results and psychometric functions for participant number 17. See Appendix E overview for full details and legend.

Table 8-22 - Experiment 2. Individual psychometric function parameter estimates for participant number 17. See Appendix E overview for full details.

Stimulus	Distance (m)	Session	Location (dB SPL)	LocationSE (dB)	Slope (ß)	SlopeSE (ß)	Deviation	pDev
PineCone	25	1	43.70	262.54	0.09	0.05	3.42	0.57
		2	47.70	2.01	6.95	0.39	0.00	0.00
	50	1	56.65	682.44	0.10	0.54	4.52	0.32
		2	59.31	2.11	0.35	1.64	2.32	0.31
	100	1	58.62	97.52	0.11	0.04	2.12	0.77
		2	67.90	1.91	0.38	1.39	3.69	0.09
Noise	25	1	43.86	9.55	0.13	0.13	0.75	0.99
		2	61.00	389.93	0.20	0.97	1.26	0.42
	50	1	62.30	93.08	0.10	0.05	3.50	0.58
		2	70.08	1.65	0.35	1.32	0.63	0.40
	100	1	68.85	2.34	0.17	0.05	5.10	0.35
		2	75.21	3.70	0.15	0.21	8.02	0.11
Whisper	25	1	4712.12	17230.62	0.00	0.08	9.21	0.11
		2	82.30	0.00	1.50	0.00	0.00	0.00
	50	1	115.84	3016.84	0.04	0.30	2.87	0.74
		2	73.90	1.70	6.95	0.56	0.00	0.01
	100	1	101.87	1815.53	0.07	0.37	3.05	0.69
		2	95.50	0.00	2.50	0.00	0.00	0.00

8.2.1.18 *Participant 18*

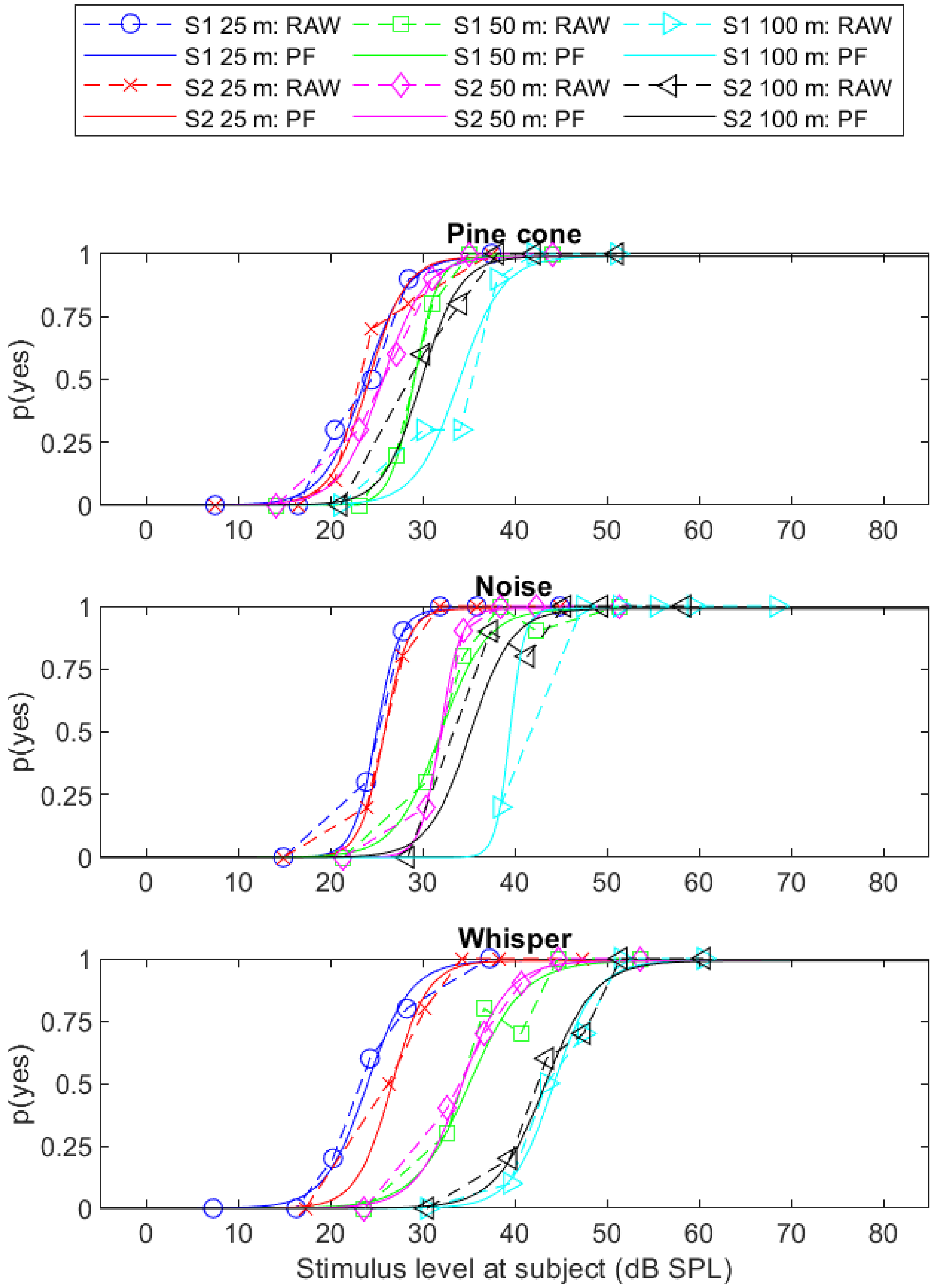

Figure 8.24 - Experiment 2. Individual results and psychometric functions for participant number 18. See Appendix E overview for full details and legend.

Table 8-23 - Experiment 2. Individual psychometric function parameter estimates for participant number 18. See Appendix E overview for full details.

Stimulus	Distance (m)	Session	Location (dB SPL)	LocationSE (dB)	Slope (ß)	SlopeSE (ß)	Deviation	pDev
PineCone	25	1	28.11	0.90	0.51	1.42	1.03	0.41
		2	29.28	1.39	0.34	0.87	1.79	0.57
	50	1	34.96	1.19	0.42	1.22	1.75	0.43
		2	34.47	0.80	0.70	1.66	0.17	0.34
	100	1	42.33	1.31	0.30	0.15	1.57	0.76
		2	38.82	1.19	0.33	0.17	1.44	0.74
Noise	25	1	30.33	1.06	0.37	0.12	0.28	0.95
		2	32.35	0.84	0.58	0.70	0.35	0.70
	50	1	37.63	0.70	0.79	0.96	0.72	0.37
		2	43.77	1.66	0.26	0.36	2.85	0.55
	100	1	48.40	1.44	0.27	0.15	3.42	0.47
		2	50.06	0.94	0.50	0.36	3.14	0.17
Whisper	25	1	33.83	1.46	0.23	0.08	2.60	0.68
		2	37.99	1.37	0.34	0.50	3.46	0.36
	50	1	43.25	1.15	0.39	0.33	1.20	0.56
		2	46.91	1.15	0.34	0.15	2.97	0.42
	100	1	53.55	1.42	0.26	0.09	2.31	0.64
		2	54.26	0.93	0.51	0.44	4.96	0.10

8.2.1.19 *Participant 19*

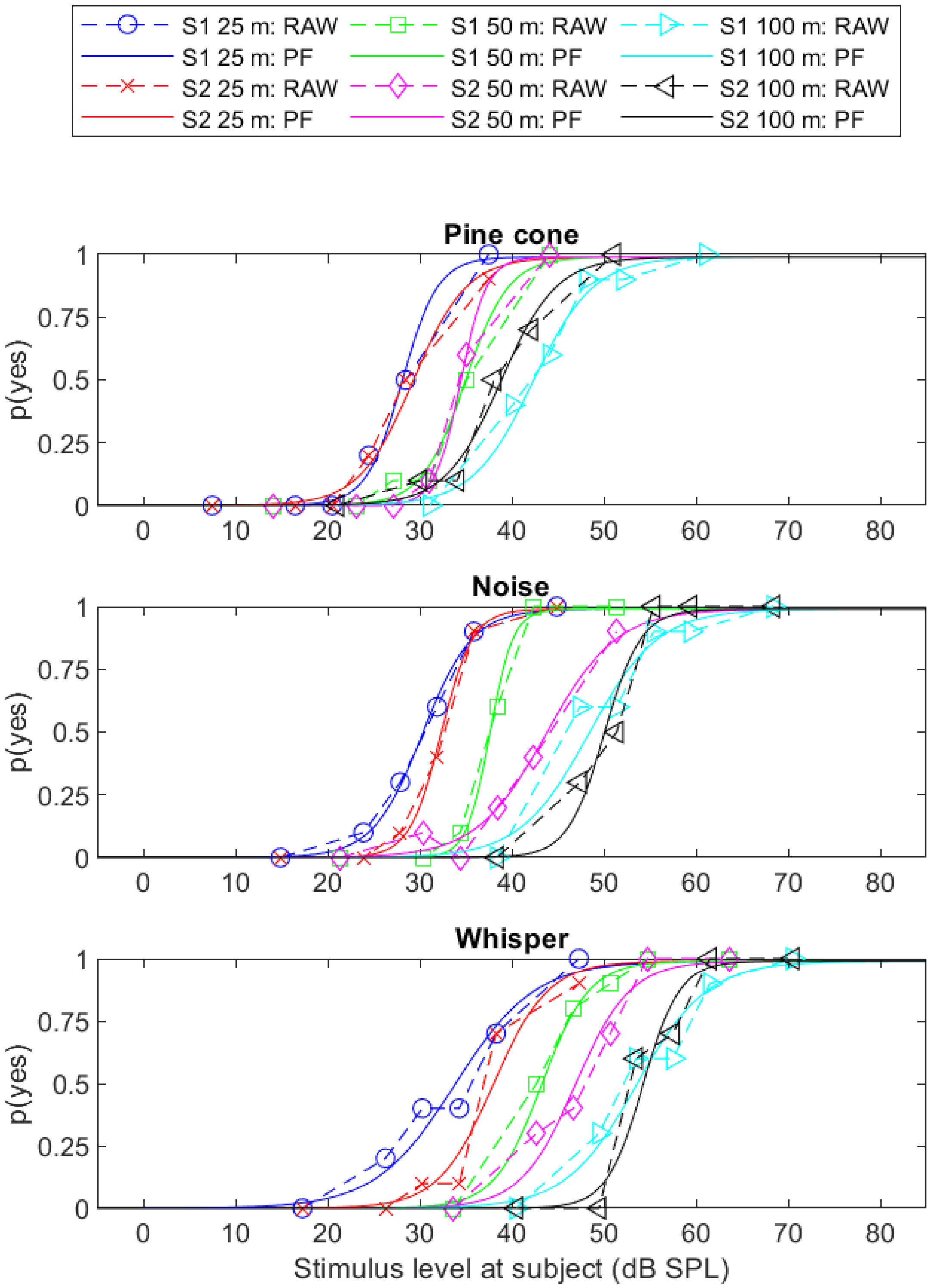

Figure 8.25 - Experiment 2. Individual results and psychometric functions for participant number 19 . See Appendix E overview for full details and legend.

Table 8-24 - Experiment 2. Individual psychometric function parameter estimates for participant number 19. See Appendix E overview for full details.

Stimulus	Distance (m)	Session	Location (dB SPL)	LocationSE (dB)	Slope (ß)	SlopeSE (ß)	Deviation	pDev
PineCone	25	1	28.11	0.90	0.51	1.42	1.03	0.41
		2	29.28	1.39	0.34	0.87	1.79	0.57
	50	1	34.96	1.19	0.42	1.22	1.75	0.43
		2	34.47	0.80	0.70	1.66	0.17	0.34
	100	1	42.33	1.31	0.30	0.15	1.57	0.76
		2	38.82	1.19	0.33	0.17	1.44	0.74
Noise	25	1	30.33	1.06	0.37	0.12	0.28	0.95
		2	32.35	0.84	0.58	0.70	0.35	0.70
	50	1	37.63	0.70	0.79	0.96	0.72	0.37
		2	43.77	1.66	0.26	0.36	2.85	0.55
	100	1	48.40	1.44	0.27	0.15	3.42	0.47
		2	50.06	0.94	0.50	0.36	3.14	0.17
Whisper	25	1	33.83	1.46	0.23	0.08	2.60	0.68
		2	37.99	1.37	0.34	0.50	3.46	0.36
	50	1	43.25	1.15	0.39	0.33	1.20	0.56
		2	46.91	1.15	0.34	0.15	2.97	0.42
	100	1	53.55	1.42	0.26	0.09	2.31	0.64
		2	54.26	0.93	0.51	0.44	4.96	0.10

8.2.1.20 *Participant 20*

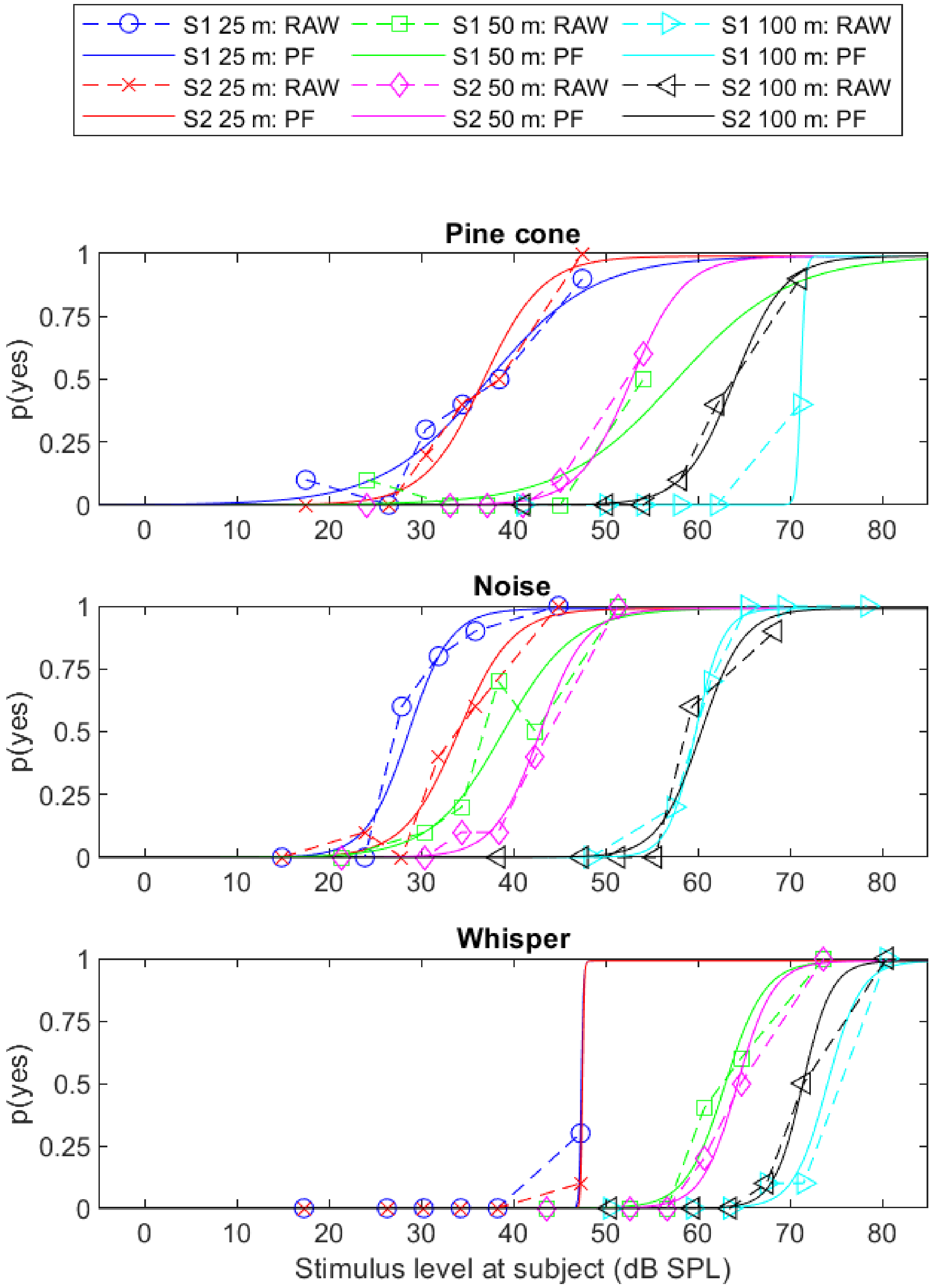

Figure 8.26 - Experiment 2. Individual results and psychometric functions for participant number 20. See Appendix E overview for full details and legend.

Table 8-25 - Experiment 2. Individual psychometric function parameter estimates for participant number 20. See Appendix E overview for full details.

Stimulus	Distance (m)	Session	Location (dB SPL)	LocationSE (dB)	Slope (ß)	SlopeSE (ß)	Deviation	pDev
PineCone	25	1	37.21	2.12	0.17	0.06	4.39	0.33
		2	36.75	1.31	0.31	0.28	2.93	0.49
	50	1	57.62	282.63	0.16	0.70	10.13	0.00
		2	52.78	1.85	0.35	1.51	0.63	0.40
	100	1	71.20	1.08	4.11	0.08	0.00	0.10
		2	63.98	1.58	0.37	1.12	0.98	0.67
Noise	25	1	28.64	1.05	0.44	0.36	4.42	0.16
		2	34.02	1.27	0.34	0.32	4.24	0.25
	50	1	38.85	1.37	0.26	0.15	5.26	0.27
		2	42.89	1.18	0.40	1.01	2.01	0.44
	100	1	59.75	0.84	0.67	0.49	0.68	0.37
		2	60.34	1.20	0.43	1.42	5.42	0.06
Whisper	25	1	47.40	0.77	8.56	0.71	0.00	0.72
		2	47.50	1.55	11.05	0.24	0.00	0.00
	50	1	62.98	1.00	0.47	0.63	2.69	0.21
		2	64.31	0.99	0.51	1.40	1.03	0.34
	100	1	74.06	2.09	0.55	2.18	2.58	0.13
		2	71.39	0.77	0.63	1.72	0.28	0.36

8.2.1.21 *Participant 21*

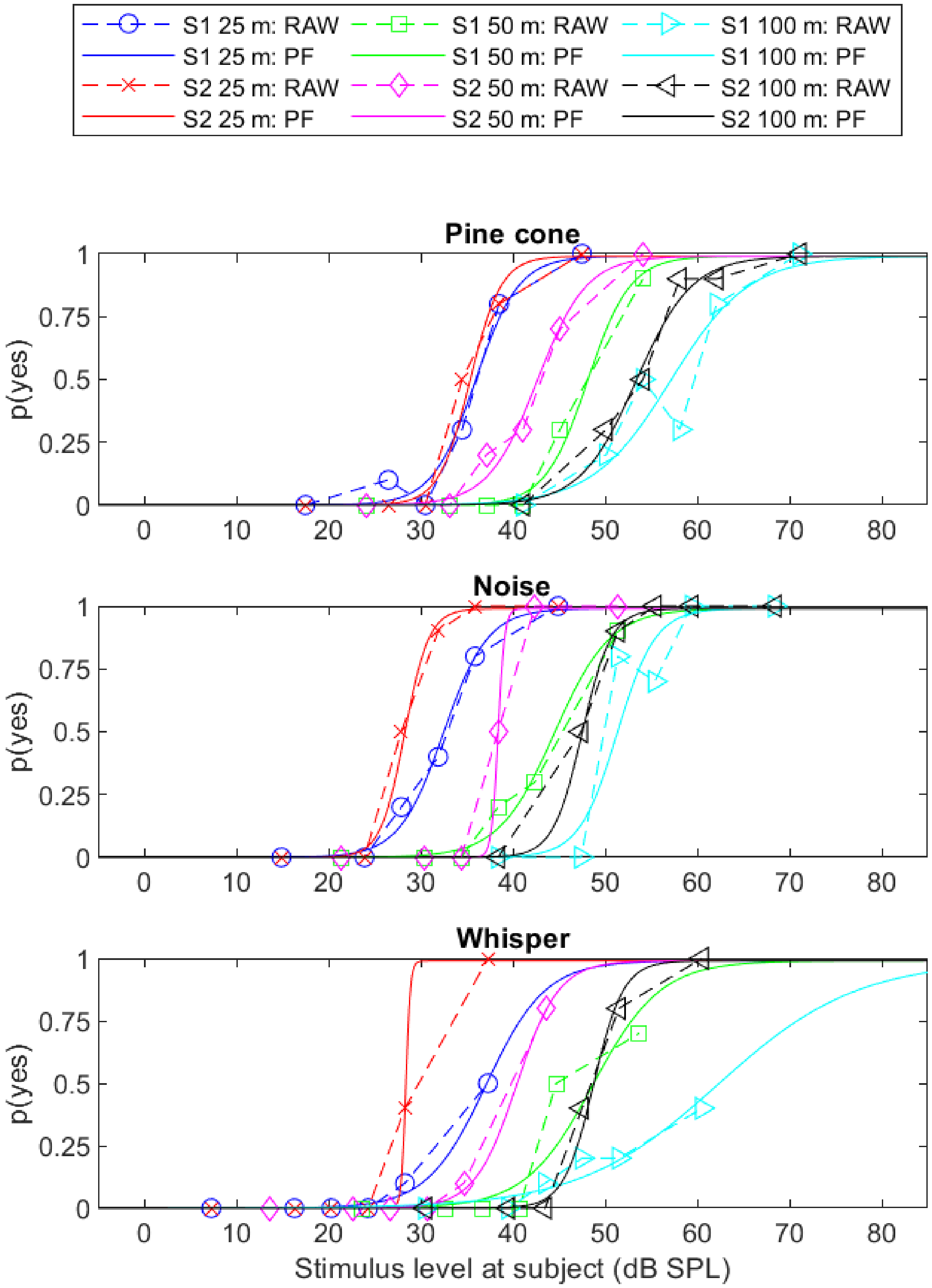

Figure 8.27 - Experiment 2. Individual results and psychometric functions for participant number 21. See Appendix E overview for full details and legend.

Table 8-26 - Experiment 2. Individual psychometric function parameter estimates for participant number 21. See Appendix E overview for full details.

Stimulus	Distance (m)	Session	Location (dB SPL)	LocationSE (dB)	Slope (ß)	SlopeSE (ß)	Deviation	pDev
PineCone	25	1	35.94	1.00	0.43	0.44	4.19	0.18
		2	35.35	0.85	0.59	0.86	2.18	0.21
	50	1	48.11	1.90	0.43	1.82	1.67	0.30
		2	42.65	1.14	0.37	0.38	1.67	0.57
	100	1	57.11	1.48	0.24	0.07	5.31	0.25
		2	53.38	1.17	0.33	0.15	1.44	0.69
Noise	25	1	32.47	1.04	0.42	0.38	1.14	0.66
		2	28.28	0.68	0.76	0.72	1.21	0.28
	50	1	44.51	1.68	0.33	1.08	1.51	0.59
		2	38.40	0.24	2.83	0.11	0.00	0.56
	100	1	51.32	0.95	0.48	0.41	9.09	0.01
		2	47.51	0.87	0.63	0.59	0.28	0.35
Whisper	25	1	37.11	2.26	0.31	1.47	0.75	0.44
		2	28.40	0.48	4.13	0.42	0.00	0.32
	50	1	48.69	2.06	0.27	0.72	6.51	0.05
		2	40.42	2.35	0.44	2.41	0.42	0.40
	100	1	62.24	445.05	0.13	0.06	1.86	0.81
		2	48.76	0.92	0.62	0.87	1.31	0.26

8.2.1.22 *Participant 22*

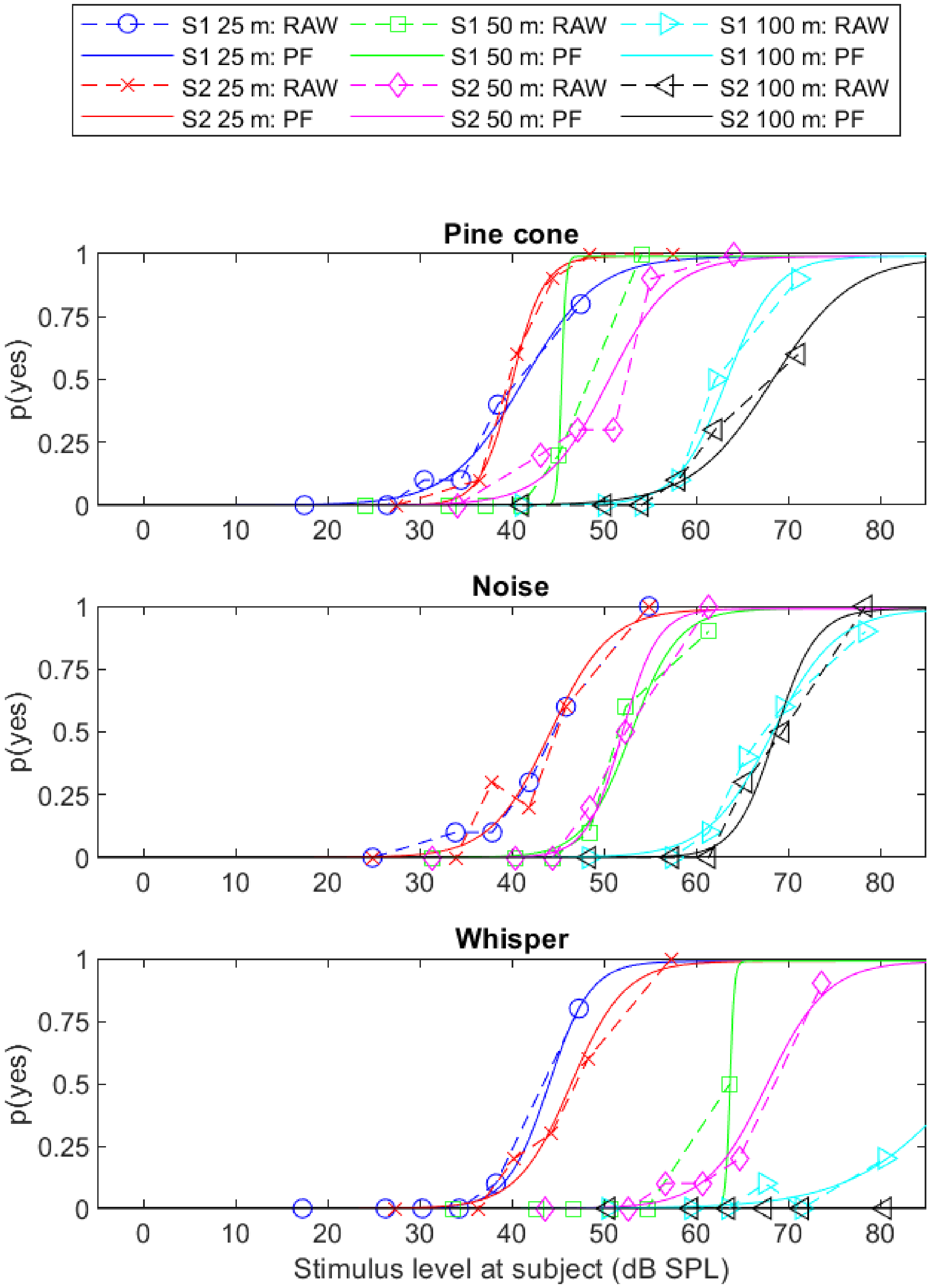

Figure 8.28 - Experiment 2. Individual results and psychometric functions for participant number 22. See Appendix E overview for full details and legend.

Table 8-27 - Experiment 2. Individual psychometric function parameter estimates for participant number 22. See Appendix E overview for full details.

Stimulus	Distance (m)	Session	Location (dB SPL)	LocationSE (dB)	Slope (ß)	SlopeSE (ß)	Deviation	pDev
PineCone	25	1	41.28	1.95	0.25	0.40	1.24	0.79
		2	40.05	0.81	0.58	0.53	0.35	0.75
	50	1	45.43	1.39	4.24	1.32	0.00	0.31
		2	50.51	1.27	0.29	0.10	4.64	0.24
	100	1	63.31	1.35	0.37	1.18	1.73	0.51
		2	68.26	2.83	0.23	0.37	1.88	0.60
Noise	25	1	44.15	1.33	0.31	0.27	1.58	0.74
		2	44.15	1.29	0.31	0.25	4.58	0.23
	50	1	52.95	1.30	0.39	0.87	2.80	0.29
		2	52.01	0.94	0.51	1.32	1.03	0.37
	100	1	68.30	1.44	0.28	0.17	1.78	0.73
		2	68.65	1.05	0.46	0.89	2.13	0.28
Whisper	25	1	44.02	2.11	0.44	2.76	0.42	0.37
		2	46.42	1.16	0.34	0.25	1.98	0.60
	50	1	63.70	0.16	4.21	0.13	0.00	0.53
		2	67.70	1.78	0.30	0.84	1.64	0.59
	100	1	89.12	467.26	0.16	0.93	2.73	0.38
		2	95.50	0.00	2.50	0.00	0.00	0.00

8.2.1.23 *Participant 23*

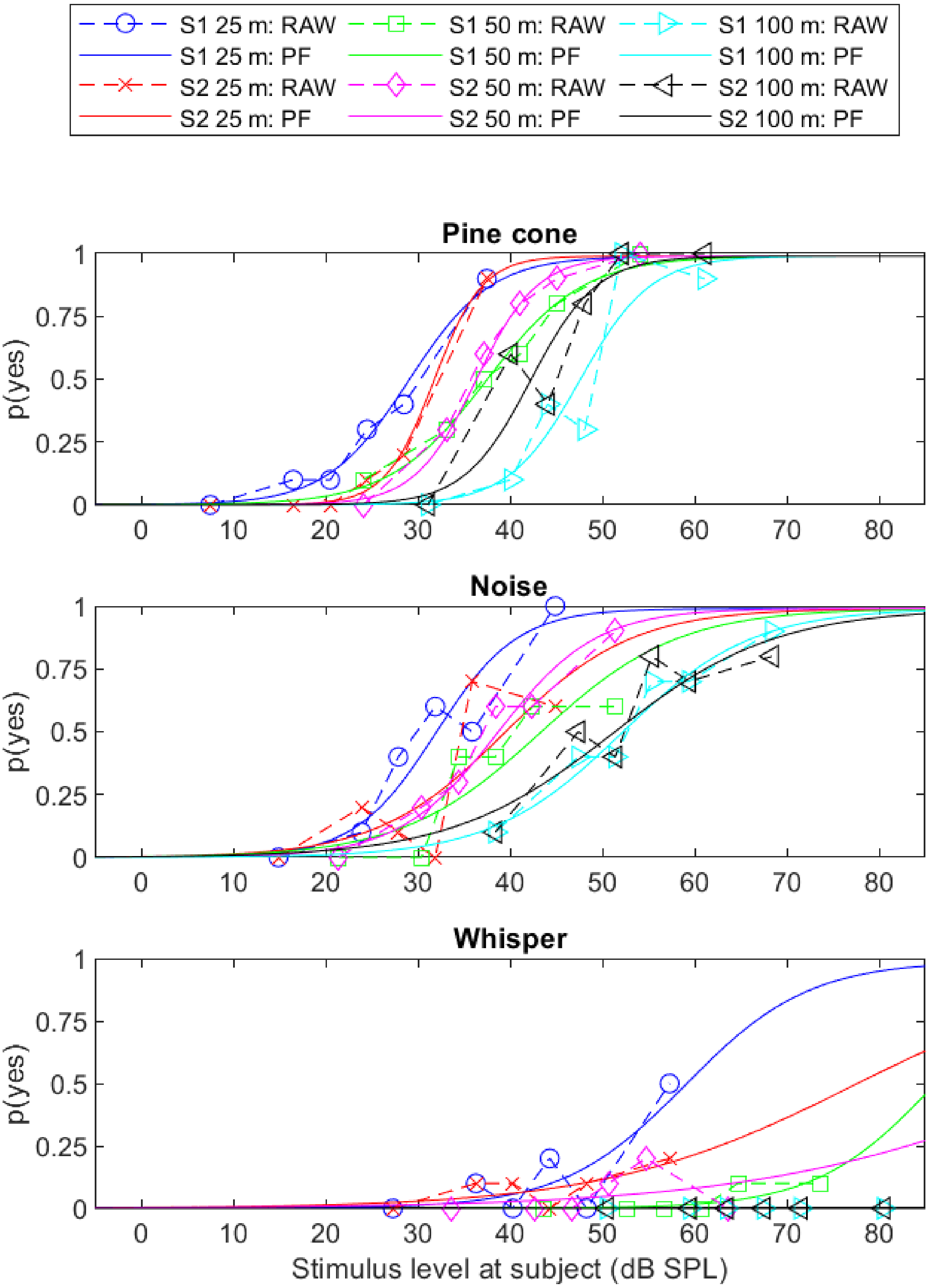

Figure 8.29 - Experiment 2. Individual results and psychometric functions for participant number 23. See Appendix E overview for full details and legend.

Table 8-28 - Experiment 2. Individual psychometric function parameter estimates for participant number 23. See Appendix E overview for full details.

Stimulus	Distance (m)	Session	Location (dB SPL)	LocationSE (dB)	Slope (ß)	SlopeSE (ß)	Deviation	pDev
PineCone	25	1	29.02	1.82	0.22	0.08	0.87	0.95
		2	31.76	1.94	0.38	1.55	0.62	0.62
	50	1	37.44	1.74	0.19	0.06	1.14	0.96
		2	36.30	1.33	0.30	0.11	0.86	0.88
	100	1	47.56	1.36	0.26	0.15	9.64	0.05
		2	42.22	1.39	0.28	0.10	6.89	0.10
Noise	25	1	31.96	1.40	0.23	0.15	4.56	0.39
		2	39.39	3.41	0.15	0.05	12.33	0.02
	50	1	43.25	2.90	0.13	0.04	7.56	0.11
		2	38.63	1.76	0.19	0.07	1.56	0.89
	100	1	52.11	2.12	0.14	0.05	0.85	0.95
		2	51.20	2.62	0.11	0.04	3.55	0.53
Whisper	25	1	59.03	7.19	0.15	0.27	7.04	0.08
		2	77.40	1620.74	0.07	0.41	2.91	0.70
	50	1	86.02	424.97	0.16	0.86	1.65	0.35
		2	101.83	2427.13	0.06	0.55	6.52	0.14
	100	1	95.50	0.00	2.50	0.00	0.00	0.00
		2	95.50	0.00	2.50	0.00	0.00	0.00

8.2.1.24 *Participant 24*

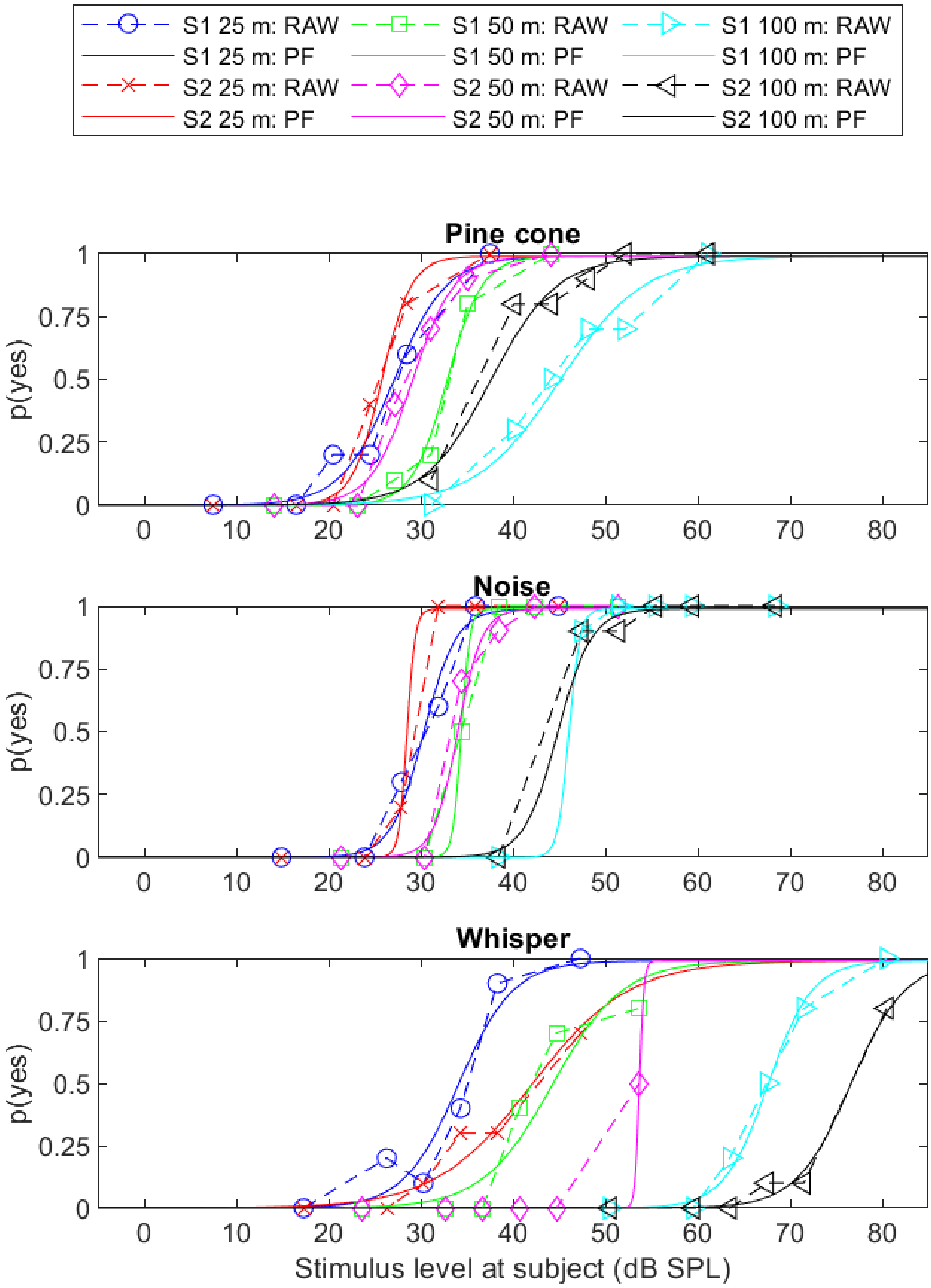

Figure 8.30 - Experiment 2. Individual results and psychometric functions for participant number 24. See Appendix E overview for full details and legend.

Table 8-29 - Experiment 2. Individual psychometric function parameter estimates for participant number 24. See Appendix E overview for full details.

Stimulus	Distance (m)	Session	Location (dB SPL)	LocationSE (dB)	Slope (ß)	SlopeSE (ß)	Deviation	pDev
PineCone	25	1	27.08	1.33	0.35	0.28	2.56	0.48
		2	25.76	0.90	0.62	0.87	1.31	0.26
	50	1	32.85	0.96	0.53	0.72	1.12	0.46
		2	29.10	0.98	0.44	0.28	2.11	0.44
	100	1	45.06	1.58	0.22	0.08	2.71	0.67
		2	37.57	1.73	0.28	0.24	2.05	0.62
Noise	25	1	30.31	0.88	0.56	0.58	2.44	0.24
		2	28.52	0.17	2.24	0.07	0.01	0.36
	50	1	34.40	0.23	2.16	0.14	0.01	0.57
		2	34.00	0.70	0.76	0.69	2.83	0.10
	100	1	46.06	0.29	1.64	0.02	0.00	0.60
		2	44.84	1.61	0.55	0.58	2.58	0.08
Whisper	25	1	34.09	1.23	0.34	0.26	4.39	0.28
		2	42.04	2.54	0.19	0.07	1.99	0.74
	50	1	44.28	1.49	0.26	0.38	6.98	0.12
		2	53.70	0.56	4.21	0.10	0.00	0.56
	100	1	67.63	0.98	0.42	0.50	1.03	0.73
		2	76.59	2.03	0.34	1.38	1.09	0.49

8.2.1.25 *Participant 25*

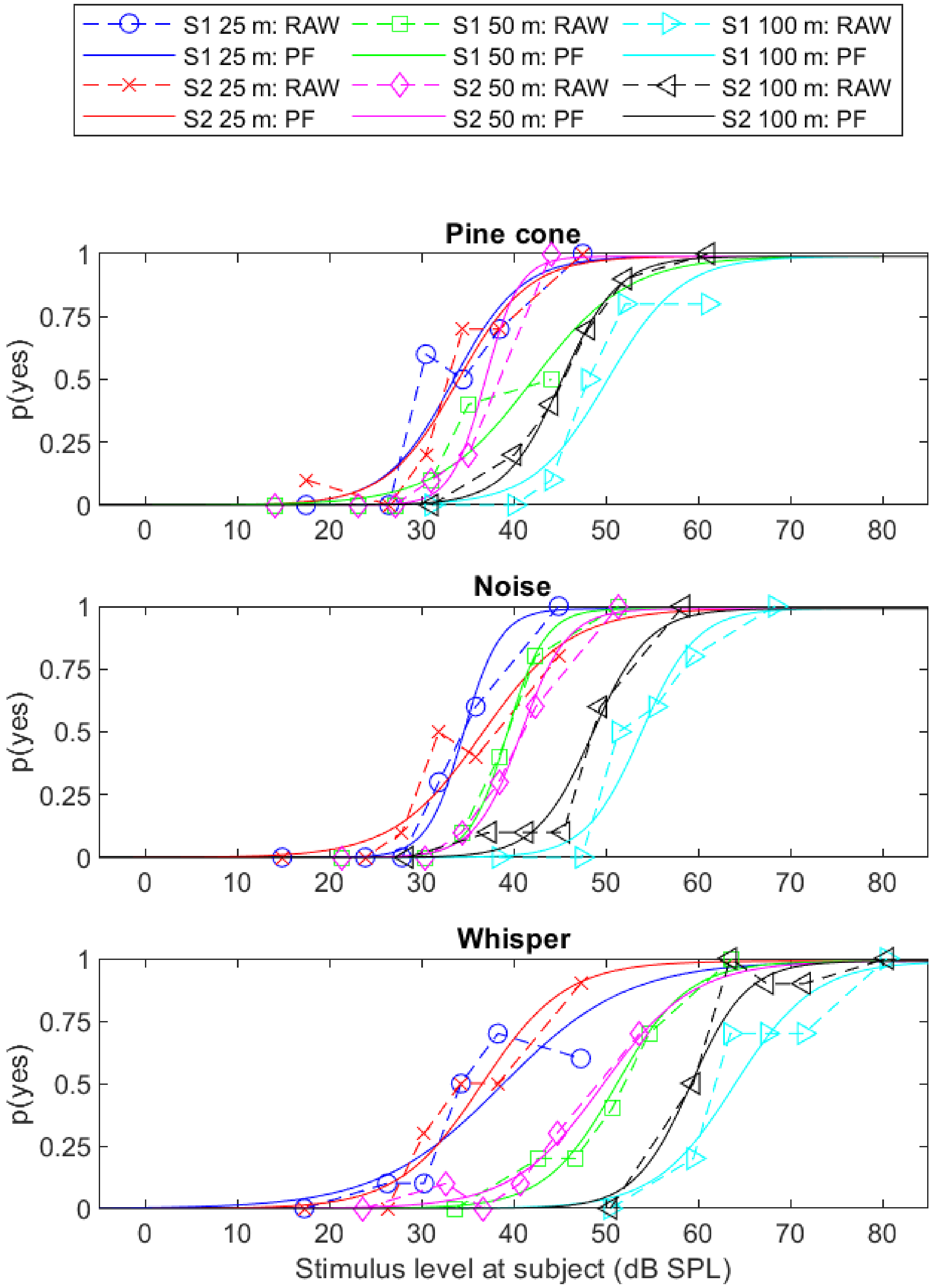

Figure 8.31 - Experiment 2. Individual results and psychometric functions for participant number 25. See Appendix E overview for full details and legend.

Table 8-30 - Experiment 2. Individual psychometric function parameter estimates for participant number 25. See Appendix E overview for full details.

Stimulus	Distance (m)	Session	Location (dB SPL)	LocationSE (dB)	Slope (ß)	SlopeSE (ß)	Deviation	pDev
PineCone	25	1	33.46	1.34	0.27	0.09	7.82	0.09
		2	33.96	1.42	0.26	0.16	7.37	0.10
	50	1	42.04	3.41	0.19	0.55	4.00	0.32
		2	36.92	1.63	0.54	2.00	1.43	0.19
	100	1	50.14	1.50	0.25	0.10	6.60	0.15
		2	45.20	1.19	0.32	0.14	0.45	0.96
Noise	25	1	34.57	0.94	0.50	0.91	1.53	0.35
		2	36.63	1.93	0.21	0.28	4.38	0.38
	50	1	39.32	1.00	0.49	0.49	0.38	0.80
		2	40.83	1.16	0.40	0.49	0.87	0.77
	100	1	53.85	1.24	0.33	0.17	4.44	0.23
		2	48.59	1.37	0.33	0.29	3.83	0.28
Whisper	25	1	38.78	2.66	0.15	0.05	6.60	0.19
		2	36.53	1.72	0.22	0.15	3.82	0.44
	50	1	51.15	1.37	0.26	0.16	2.15	0.70
		2	49.68	2.50	0.20	0.26	2.70	0.61
	100	1	63.97	1.44	0.23	0.09	5.09	0.32
		2	59.22	1.48	0.33	0.21	7.66	0.05

8.2.1.26 *Participant 26*

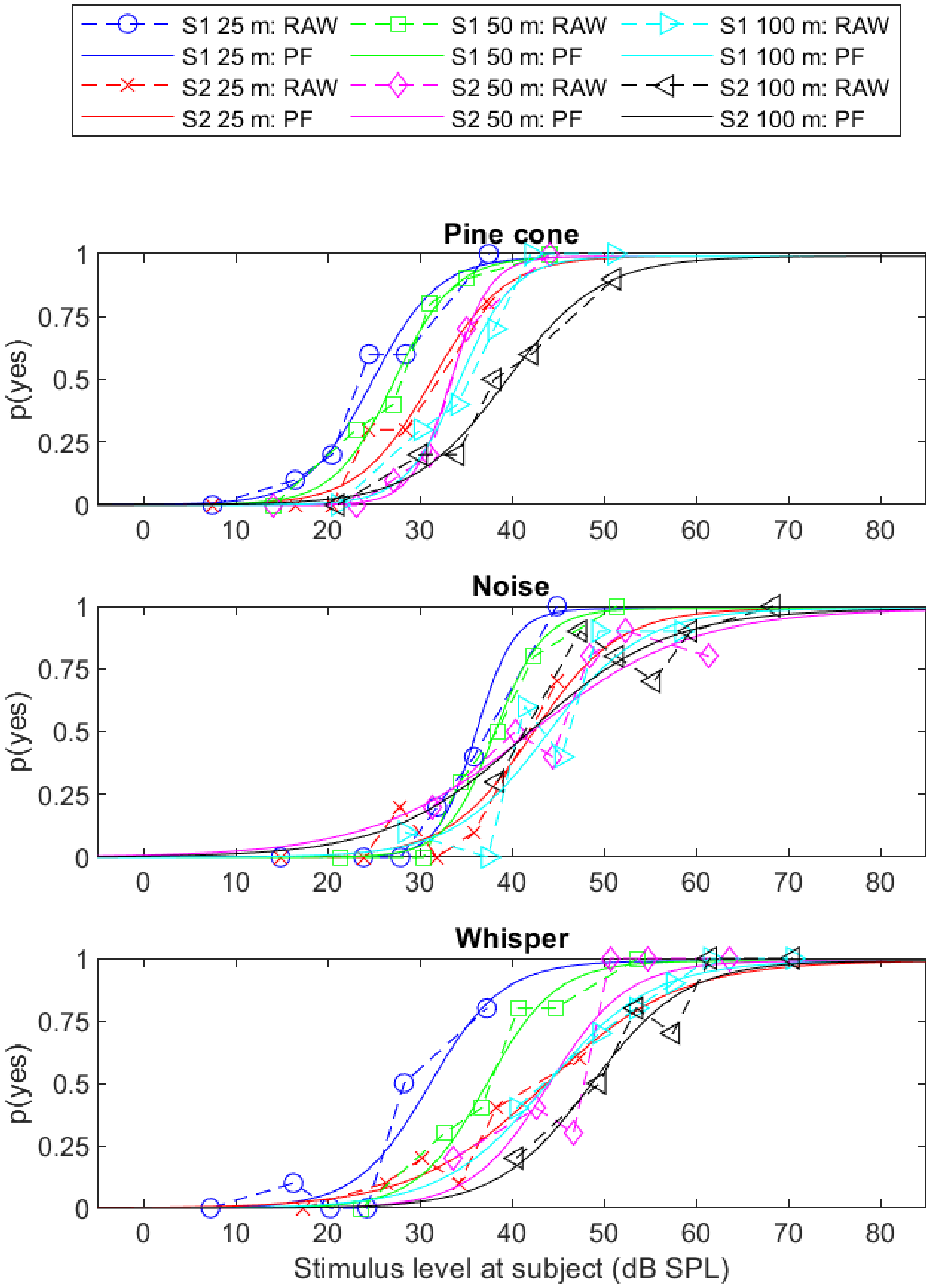

Figure 8.32 - Experiment 2. Individual psychometric function parameter estimates for participant number 26. See Appendix E overview for full details.

Table 8-31 - Experiment 2. Individual psychometric function parameter estimates for participant number 26. See Appendix E overview for full details.

Stimulus	Distance (m)	Session	Location (dB SPL)	LocationSE (dB)	Slope (ß)	SlopeSE (ß)	Deviation	pDev
PineCone	25	1	24.89	1.30	0.28	0.10	2.27	0.63
		2	31.28	1.91	0.25	0.29	3.27	0.42
	50	1	27.26	1.25	0.31	0.14	1.25	0.82
		2	33.38	1.04	0.47	0.77	0.88	0.60
	100	1	34.31	1.18	0.34	0.25	2.97	0.41
		2	39.41	1.79	0.20	0.07	1.24	0.94
Noise	25	1	36.12	1.04	0.48	1.50	1.53	0.40
		2	42.30	2.64	0.21	0.28	6.66	0.06
	50	1	38.14	1.11	0.38	0.24	1.93	0.51
		2	41.98	3.20	0.11	0.04	4.17	0.48
	100	1	43.95	1.63	0.19	0.06	10.45	0.03
		2	41.76	3.98	0.13	0.05	5.38	0.29
Whisper	25	1	31.08	2.04	0.25	0.56	7.34	0.06
		2	43.65	4.70	0.14	0.05	2.08	0.75
	50	1	37.25	1.39	0.27	0.16	2.10	0.69
		2	44.02	1.65	0.24	0.08	12.09	0.01
	100	1	43.83	3.43	0.18	0.10	1.37	0.84
		2	48.60	2.24	0.20	0.06	3.73	0.44

8.2.1.27 *Participant 27*

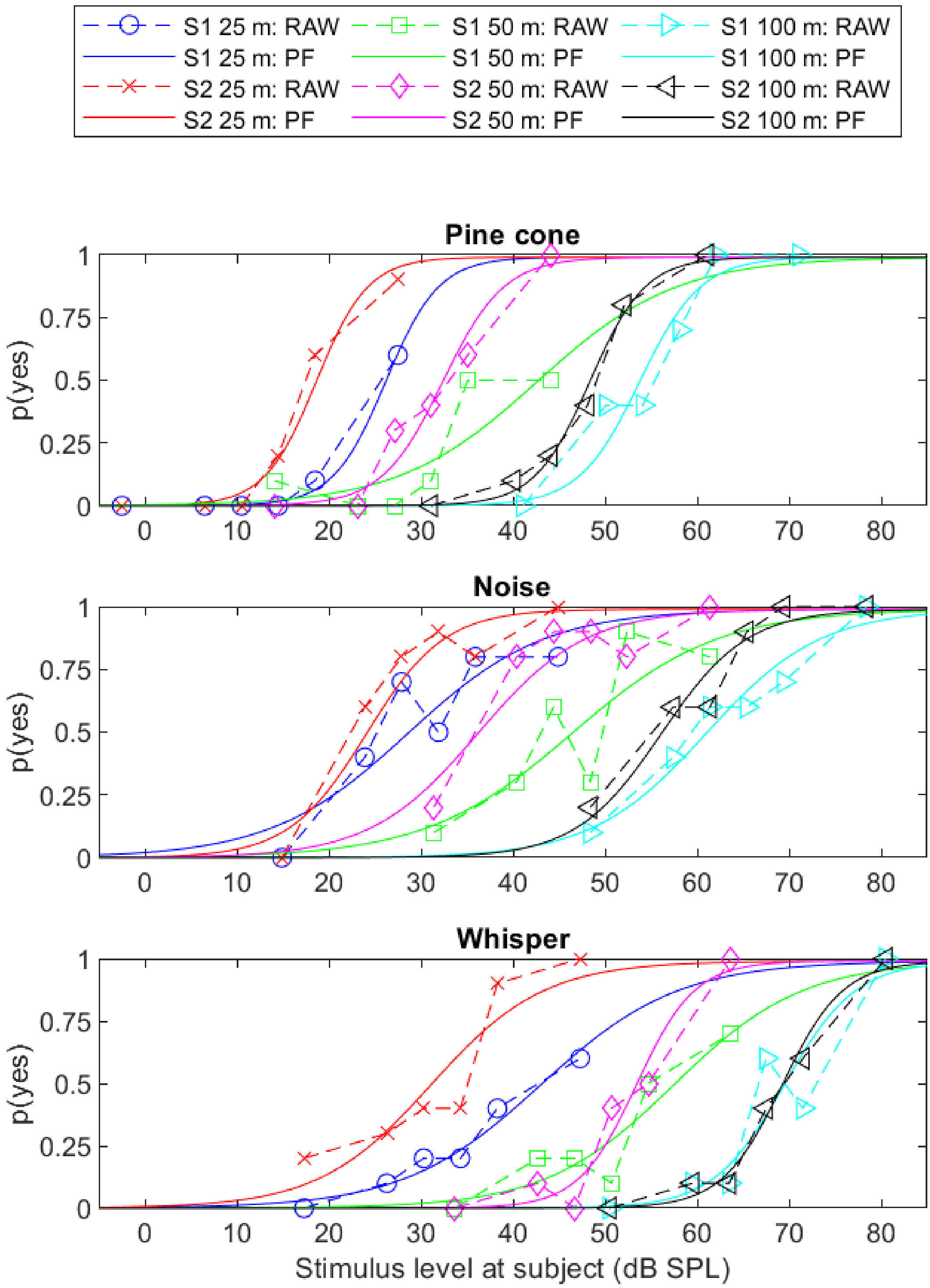

Figure 8.33 - Experiment 2. Individual results and psychometric functions for participant number 27. See Appendix E overview for full details and legend.

Table 8-32 - Experiment 2. Individual psychometric function parameter estimates for participant number 27. See Appendix E overview for full details.

Stimulus	Distance (m)	Session	Location (dB SPL)	LocationSE (dB)	Slope (ß)	SlopeSE (ß)	Deviation	pDev
PineCone	25	1	26.18	1.87	0.35	0.14	0.63	0.38
		2	18.63	1.90	0.36	0.77	2.51	0.44
	50	1	42.66	2.89	0.13	0.83	8.33	0.09
		2	32.39	1.30	0.30	0.57	3.28	0.39
	100	1	53.78	1.27	0.30	0.85	4.25	0.26
		2	48.36	2.67	0.33	0.80	0.87	0.85
Noise	25	1	28.58	1.09	0.13	0.23	6.76	0.23
		2	24.10	2.23	0.24	0.80	5.69	0.19
	50	1	46.40	2.70	0.13	0.84	7.06	0.16
		2	36.19	2.30	0.17	0.48	4.65	0.32
	100	1	60.99	2.23	0.16	0.28	2.27	0.79
		2	56.14	1.90	0.21	0.80	2.67	0.61
Whisper	25	1	43.04	1.40	0.13	0.26	1.05	0.94
		2	31.11	1.04	0.16	0.24	5.92	0.31
	50	1	57.50	2.91	0.14	0.40	3.69	0.53
		2	53.42	1.23	0.32	0.26	4.88	0.18
	100	1	69.49	1.06	0.26	0.11	6.04	0.21
		2	69.27	1.85	0.30	0.98	1.57	0.77

8.2.1.28 *Participant 28*

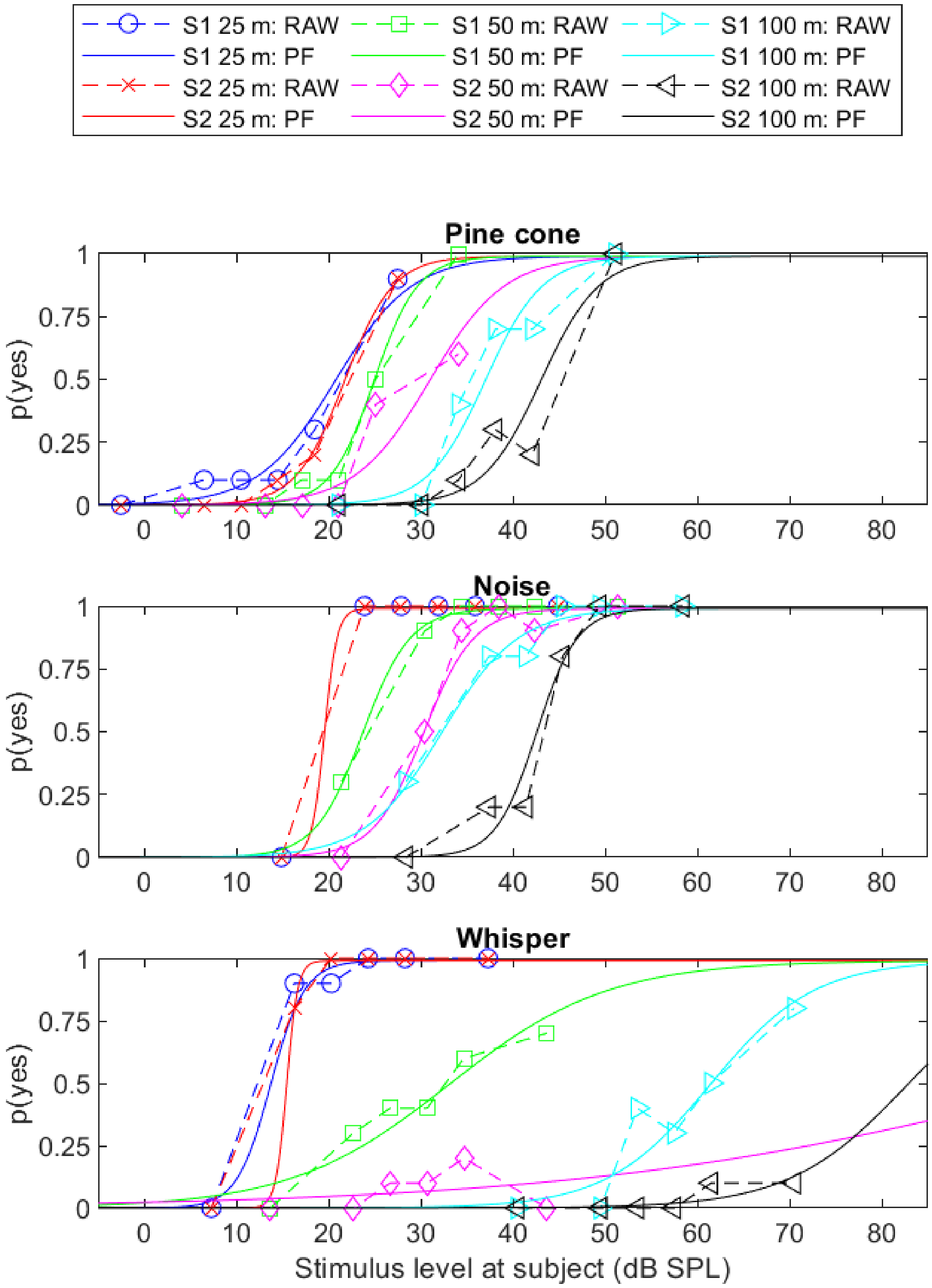

Figure 8.34 - Experiment 2. Individual results and psychometric functions for participant number 28. See Appendix E overview for full details and legend.

Table 8-33 - Experiment 2. Individual psychometric function parameter estimates for participant number 28. See Appendix E overview for full details.

Stimulus	Distance (m)	Session	Location (dB SPL)	LocationSE (dB)	Slope (ß)	SlopeSE (ß)	Deviation	pDev
PineCone	25	1	20.74	2.02	0.24	0.28	2.20	0.61
		2	21.76	2.54	0.38	0.69	0.62	0.70
	50	1	24.96	2.62	0.42	0.13	1.75	0.44
		2	30.97	1.64	0.25	0.51	5.25	0.17
	100	1	37.00	2.27	0.32	0.53	4.75	0.21
		2	43.04	2.95	0.32	0.02	5.41	0.17
Noise	25	1	19.56	1.20	1.45	0.52	0.06	0.00
		2	19.56	2.12	1.45	0.47	0.06	0.00
	50	1	23.72	2.85	0.39	0.87	0.53	0.39
		2	30.32	2.48	0.37	0.38	4.89	0.15
	100	1	32.26	1.81	0.24	0.36	2.13	0.54
		2	42.56	1.29	0.45	0.90	3.52	0.19
Whisper	25	1	13.74	1.57	0.55	0.68	2.58	0.12
		2	15.48	2.39	1.68	0.79	0.01	0.41
	50	1	33.09	1.75	0.12	0.83	2.97	0.60
		2	101.40	1.38	0.04	0.23	5.94	0.23
	100	1	61.41	1.03	0.18	0.66	5.06	0.31
		2	82.82	2.42	0.16	0.32	1.65	0.37

List of References

ABEL, S. M. 2008. Barriers to hearing conservation programs in combat arms occupations. *Aviation, space, and environmental medicine,* 79**,** 591-598.

ARMBRÜSTER, C., WOLTER, M., KUHLEN, T., SPIJKERS, W. & FIMM, B. 2008. Depth perception in virtual reality: distance estimations in peri-and extrapersonal space. *Cyberpsychology & Behavior,* 11**,** 9-15.

ATHALYE, S. P. 2010. *Factors Affecting Speech Recognition in Noise and Hearing Loss in Adults With a Wide Variety of Auditory Capabilities.* PhD Thesis, University of Southampton.

BAER, T. & MOORE, B. C. 1994. Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech. *The Journal of the Acoustical Society of America,* 95**,** 2277-2280.

BAER, T. & MOORE, B. C. J. 1993. Effects of spectral smearing on the intelligibility of sentences in noise. *The Journal of the Acoustical Society of America,* 94**,** 1229-1241.

BARALDI, A. N. & ENDERS, C. K. 2010. An introduction to modern missing data analyses. *Journal of School Psychology,* 48**,** 5-37.

BARRETT, H. C. 2016. Adaptations to predators and prey. *In:* BUSS, D. M. (ed.) *The handbook of evolutionary psychology.* New Jersey: Wiley.

BASS, H. E., SUTHERLAND, L. C., ZUCKERWAR, A. J., BLACKSTOCK, D. T. & HESTER, D. M. 1995. Atmospheric absorption of sound: Further developments. *The Journal of the Acoustical Society of America,* 97**,** 680-683.

BEE, L. 1991. Effect of magnification on distance estimation. *Army personnel research establishment ministry of defence, Farnborough Hampshire*.

BERMÚDEZ-CUAMATZIN, E., RÍOS-CHELÉN, A. A., GIL, D. & GARCIA, C. M. 2009. Strategies of song adaptation to urban noise in the house finch: syllable pitch plasticity or differential syllable use? *Behaviour***,** 1269-1286.

BEVIS, Z. L., SEMERARO, H. D., VAN BESOUW, R. M., ROWAN, D., LINETON, B. & ALLSOPP, A. J. 2014. Fit for the frontline? A focus group exploration of auditory tasks carried out by infantry and combat support personnel. *Noise and Health,* 16**,** 127.

BIGGS, T. & EVEREST, A. 2011. British military hearing conservation programme. *Clinical Otolaryngology,* 36**,** 299-301.

BILZON, J. L., SCARPELLO, E. G., BILZON, E. & ALLSOPP, A. J. 2002. Generic task-related occupational requirements for Royal Naval personnel. *Occupational Medicine,* 52**,** 503-510.

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

BLAUERT, J. 1997. *Spatial hearing: the psychophysics of human sound localization*, MIT press.

BOHIL, C. J., ALICEA, B. & BIOCCA, F. A. 2011. Virtual reality in neuroscience research and therapy. *Nature reviews neuroscience,* 12**,** 752.

BONCORAGLIO, G. & SAINO, N. 2007. Habitat structure and the evolution of bird song: a meta-analysis of the evidence for the acoustic adaptation hypothesis. *Functional Ecology,* 21**,** 134-142.

BRIMIJOIN, W. O., BOYD, A. W. & AKEROYD, M. A. 2013. The Contribution of Head Movement to the Externalization and Internalization of Sounds. *PLOS ONE,* 8**,** e83068.

BRUNGART, D. 2014. *20Q: Auditory Fitness for Duty* [Online]. Available: http://www.audiologyonline.com/articles/20q-auditory-fitness-for-duty-12528 [Accessed 03/03/2015].

BRUNGART, D. S. Control of perceived distance in virtual audio displays. Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE, 1998. IEEE, 1101-1104.

BRUNGART, D. S. & SCOTT, K. R. 2001. The effects of production and presentation level on the auditory distance perception of speech. *The Journal of the Acoustical Society of America,* 110**,** 425-440.

BS 1993. BS EN 24869-1:1993. *Sound attenuation of hearing protectors - Part 1: Subjective method of measurement.* British Standards.

BS 2009. BS EN ISO 8253-2:2009. *Acoustics. Audiometric test methods. Sound field audiometry with pure-tone and narrow-band test signals.* BS EN ISO.

BSA. 2011. Recommended Procedure: Pure-tone air-conduction and bone-conduction threshold audiometry with and without masking. *BSA Recommended Procedures* [Online]. [Accessed 14/07/2014].

BUCK, L. E., YOUNG, M. K. & BODENHEIMER, B. 2018. A comparison of distance estimation in HMD-based virtual environments with different HMD-based conditions. *ACM Transactions on Applied Perception (TAP),* 15**,** 21.

BUUS, S., FLORENTINE, M. & POULSEN, T. 1997. Temporal integration of loudness, loudness discrimination, and the form of the loudness function. *The journal of the acoustical society of America,* 101**,** 669-680.

CARLSON, N. R., BUSKIST, W., ENZLE, M. E. & HETH, C. D. 2000. Psychology: The Science of Behaviour (Canadian Edition). Allyn and Bacon Canada (CN): Scarborough, Ontario.

CASALI, J. G., AHROON, W. A. & LANCASTER, J. A. 2009. A field investigation of hearing protection and hearing enhancement in one device: For soldiers whose ears and lives depend upon it. *Noise and Health,* 11**,** 69.

COLES, R. R. A. 1964. *Hearing of "personnel sounds". Hearing of sounds important in auditory detection of presence of enemy personnel: an account of the deleterious effects on this faculty of previous exposure to noise,* Alverstoke, UK, Royal Naval Medical School.

CREEM-REGEHR, S. H., STEFANUCCI, J. K., THOMPSON, W. B., NASH, N. & MCCARDELL, M. 2016. Egocentric Distance Perception in the Oculus Rift (DK2). *Sap 2015: Acm Siggraph Symposium on Applied Perception***,** 47-50.

CRONBACH, L. J. & MEEHL, P. E. 1955. Construct validity in psychological tests. *Psychological bulletin,* 52**,** 281.

CUTTING, J. E. & VISHTON, P. M. 1995. Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. *In:* EPSTEIN, W. & ROGERS, S. (eds.) *Handbook of perception and cognition; Perception of space and motion.* San Diego, CA: Academic Press.

DAUM, S. O. & HECHT, H. 2009. Distance estimation in vista space. *Attention, Perception, & Psychophysics,* 71**,** 1127-1137.

DAVIS, M., GENDELMAN, D., TISCHLER, M. & GENDELMAN, P. 1982. A primary acoustic startle circuit: lesion and stimulation studies. *The Journal of Neuroscience,* 2**,** 791-805.

DEFENCE, M. O. 2013a. JSP 950 Annex D to Leaflet 6-7-3. Assessment of Hearing Acuity (H).

DEFENCE, M. O. 2013b. JSP 950 Annex I to Leaflet 6-7-3. Guidelines for undertaking screening pure tone audiometry. *In:* DEFENCE, M. O. (ed.).

DEFENCE, M. O. 2013c. JSP 950: Medical Assessment for PULHEEMS. *In:* DEFENCE, M. O. (ed.).

DRULLMAN, R., FESTEN, J. M. & PLOMP, R. 1994. Effect of temporal envelope smearing on speech reception. *The Journal of the Acoustical Society of America,* 95**,** 1053-1064.

DUCHNOWSKI, P. 1989. *Simulation of sensorineural hearing impairment.* Doctoral thesis, Massachusetts Institute of Technology.

EGNOR, S. R., WICKELGREN, J. G. & HAUSER, M. D. 2007. Tracking silence: adjusting vocal production to avoid acoustic interference. *Journal of Comparative Physiology A,* 193**,** 477-483.

ENDSLEY, M. R. 1995. Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society,* 37**,** 32-64.

ENDSLEY, M. R., GARLAND, D. J., WAMPLER, R. L. & MATTHEWS, M. D. 2000. Modeling and measuring situation awareness in the infantry operational environment. US Army Research Institute for the Behavioural and Social Sciences.

ENGELBERG, M. 1965. Relationship of Pure Tones to Speech Reception Threshold. *Annals of Otology, Rhinology and Laryngology,* 74**,** 234-40.

ESQUIVEL, C. R., PARKER, M., CURTIS, K., MERKLEY, A., LITTLEFIELD, P., CONLEY, G., WISE, S., FELDT, B., HENSELMAN, L. & STOCKINGER, Z. 2018. Aural blast injury/acoustic trauma and hearing loss. *Military medicine,* 183**,** 78-82.

FABRY, D. A. & VAN TASELL, D. J. 1986. Masked and filtered simulation of hearing loss: effects on consonant recognition. *Journal of Speech and Hearing Research,* 29**,** 170-8.

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.

FELTY, R. A., BUCHWALD, A. & PISONI, D. B. 2009. Adaptation to frozen babble in spoken word recognition. *The Journal of the Acoustical Society of America,* 125**,** EL93-EL97.

FERMAN, L., VERSCHUURE, J. & VAN ZANTEN, B. 1993. Impaired speech perception in noise in patients with a normal audiogram. *Audiology,* 32**,** 49-54.

FINE, B. J. & KOBRICK, J. L. 1983. Individual differences in distance estimation: Comparison of judgments in the field with those from projected slides of the same scenes. *Perceptual and motor skills,* 57**,** 3-14.

FITZPATRICK, V., PASNAK, R. & TYER, Z. E. 1982. The effect of familiar size at familiar distances. *Perception,* 11**,** 85-91.

FLETCHER, R. T. 1949. Pulheems: System of Medical Classification. *British medical journal,* 1**,** 83.

FLORENTINE, M., BUUS, S. R. & POULSEN, T. 1996. Temporal integration of loudness as a function of level. *The Journal of the Acoustical Society of America,* 99**,** 1633-1644.

FOREMAN, N., SANDAMAS, G. & NEWSON, D. 2004. Distance underestimation in virtual space is sensitive to gender but not activity-passivity or mode of interaction. *CyberPsychology & Behavior,* 7**,** 451-457.

GARINTHER, G. R., KALB, J. T., HODGE, D. C. & PRICE, G. R. 1985. Proposed aural nondetectability limits for army materiel. HUMAN ENGINEERING LAB ABERDEEN PROVING GROUND MD.

GARINTHER, G. R., WHITAKER, L. A. & PETERS, L. J. The effects of speech intelligibility on military performance. Proceedings of the Symposium on Speech Communication Metrics and Human Performance, 1995. DTIC Document, 72-82.

GASTON, J., MERMAGEN, T. & DICKERSON, K. 2013. Acoustic Measurement and Model Predictions for the Aural Nondetectability of Two Night-Vision Goggles. ARMY RESEARCH LAB ABERDEEN PROVING GROUND MD HUMAN RESEARCH AND ENGINEERING DIRECTORATE.

GATEHOUSE, S. & HAGGARD, M. P. 1987. The effects of air-bone gap and presentation level on word identification. *Ear Hear,* 8**,** 140-6.

GESCHEIDER, G. A. 2013. *Psychophysics*, Psychology Press.

GIGUÈRE, C., LAROCHE, C., SOLI, S., LAGACÉ, J., FORTIN, M. & VAILLANCOURT, V. Hearing Standards for Fisheries & Oceans Canada. Proceedings of the 8th International Congress on Noise as Public Health Problem, Rotterdam, 2003.

GIGUERE, C., LAROCHE, C., SOLI, S. D. & VAILLANCOURT, V. 2008. Functionally-based screening criteria for hearing-critical jobs based on the Hearing in Noise Test. *International journal of audiology,* 47**,** 319-328.

GILINSKY, A. S. 1951. Perceived size and distance in visual space. *Psychological review,* 58**,** 460.

GLASBERG, B. R. & MOORE, B. C. 2002. A model of loudness applicable to time-varying sounds. *Journal of the Audio Engineering Society,* 50**,** 331-342.

GLASBERG, B. R. & MOORE, B. C. J. 2006. Prediction of absolute thresholds and equal-loudness contours using a modified loudness model. *Journal of the Acoustical Society of America,* 120**,** 585-588.

GOEHRING, T., CHAPMAN, J. L., BLEECK, S. & MONAGHAN, J. J. M. 2018. Tolerable delay for speech production and perception: effects of hearing ability and experience with hearing aids. *International Journal of Audiology,* 57**,** 61-68.

GOERLITZ, H. R., TER HOFSTEDE, H. M., ZEALE, M. R. K., JONES, G. & HOLDERIED, M. W. 2010. An Aerial-Hawking Bat Uses Stealth Echolocation to Counter Moth Hearing. *Current Biology,* 20**,** 1568-1572.

HARRIS, J. D. 1965. Pure-Tone Acuity and the Intelligibility of Everyday Speech. *Journal of the Acoustical Society of America,* 37**,** 824-30.

HART, S. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* 50**,** 904-908.

HEINRICH, A., HENSHAW, H. & FERGUSON, M. A. 2015. The relationship of speech intelligibility with hearing sensitivity, cognition, and perceived hearing difficulties varies for different speech perception tests. *Frontiers in Psychology,* 6.

HELFER, T. M., CANHAM-CHERVAK, M., CANADA, S. & MITCHENER, T. A. 2010. Epidemiology of hearing impairment and noise-induced hearing injury among US military personnel, 2003–2005. *American journal of preventive medicine,* 38**,** S71-S77.

HODGETTS, T. & GREASLEY, L. 2003. Impact Of Deployment Of Personnel With Chronic Conditions To Forward Areas. *Journal of the Royal Army Medical Corps,* 149**,** 277-283.

HOLMES, H. 2015. *The effects of background sound on communication: speech intelligibility and reading.* University of Southampton.

HOUTGAST, T. & FESTEN, J. M. 2008. On the auditory and cognitive functions that may explain an individual's elevation of the speech reception threshold in noise. *International journal of audiology,* 47**,** 287-295.

HSE 2005. Controlling noise at work. Guidance on the Control of Noise at Work Regulations 2005. *In:* EXECUTIVE, H. A. S. (ed.). HSE Books.

HUMES, L., JOELLENBECK, L. M. & DURCH, J. 2005. Noise and military service implications for hearing loss and tinnitus. Washington, DC: National Academies Press.

HUMES, L. E. & ROBERTS, L. 1990. Speech-recognition difficulties of the hearing-impaired elderly: The contributions of audibility. *Journal of Speech and Hearing Research,* 33**,** 726.

JEFFERSON, T. 1990. An investigation into regular recruit wastage from the British Army, 1988. *Journal of the Royal Army Medical Corps,* 136**,** 138-145.

JOHNSON, C. J., PICK JR, H. L., SIEGEL, G. M., CICCIARELLI, A. W. & GARBER, S. R. 1981. Effects of interpersonal distance on children's vocal intensity. *Child Development***,** 721-723.

KAHNEMAN, D. 1973. *Attention and Effort,* New Jersey, Prentice-Hall.

KELLER, M. D., ZIRIAX, J. M., BARNS, W., SHEFFIELD, B., BRUNGART, D., THOMAS, T., JAEGER, B. & YANKASKAS, K. 2017. Performance in noise: Impact of reduced speech intelligibility on Sailor performance in a Navy command and control environment. *Hearing Research,* 349**,** 55-66.

KILLION, M. C. & NIQUETTE, P. A. 2000. What can the pure-tone audiogram tell us about a patient's SNR loss? *The Hearing Journal,* 53**,** 46-48.

KLEIN, S. A. 2001. Measuring, estimating, and understanding the psychometric function: A commentary. *Perception & psychophysics,* 63**,** 1421-1455.

KNAPP, J. L. J. 2003. Visual perception of egocentric distance in real and virtual environments. *In:* HETTINGER, L. J. & HAAS, M. W. (eds.) *Virtual and adaptive environments.* New Jersey, US: CRC Press.

KNAPP, J. M. & LOOMIS, J. M. 2004. Limited field of view of head-mounted displays is not the cause of distance underestimation in virtual environments. *Presence: Teleoperators & Virtual Environments,* 13**,** 572-577.

KÖHLER, W. & WINTER, E. 1927. *The mentality of apes: Translated from the second revised edition by Ella Winter,* London, Routledge.

KOLARIK, A. J., MOORE, B. C. J., ZAHORIK, P., CIRSTEA, S. & PARDHAN, S. 2016. Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss. *Attention, Perception, & Psychophysics,* 78**,** 373-395.

KRAMER, S. E., KAPTEYN, T. S., FESTEN, J. M. & TOBI, H. 1996. The relationships between self-reported hearing disability and measures of auditory disability. *International Journal of Audiology,* 35**,** 277-287.

KÜHNAPFEL, U., CAKMAK, H. K. & MAAß, H. 2000. Endoscopic surgery training using virtual reality and deformable tissue simulation. *Computers & graphics,* 24**,** 671-682.

KUJAWA, S. G. & LIBERMAN, M. C. 2009. Adding insult to injury: cochlear nerve degeneration after "temporary" noise-induced hearing loss. *The Journal of Neuroscience,* 29**,** 14077-14085.

LAROCHE, C., SOLI, S., GIGUERE, C., LAGACÉ, J., VAILLANCOURT, V. & FORTIN, M. 2003. An approach to the development of hearing standards for hearing-critical jobs. *Noise and Health,* 6**,** 17.

LAROCHE, C., SOLI, S., GIGUERE, C. & VAILLANCOURT, V. Establishment of fitness standards for hearing-critical jobs. 9th International Congress on Noise as a Public Health Problem, 2008 Los Angeles.

LEWALD, J. 2013. Exceptional ability of blind humans to hear sound motion: Implications for the emergence of auditory space. *Neuropsychologia,* 51**,** 181-186.

LITTLE, R. J. & RUBIN, D. B. 2014. *Statistical analysis with missing data*, John Wiley & Sons.

LOYOLA, M. 2018. The influence of the availability of visual cues on the accurate perception of spatial dimensions in architectural virtual environments. *Virtual Reality,* 22**,** 235-243.

LUM, D. S. & BRAIDA, L. D. 2000. Perception of speech and non-speech sounds by listeners with real and simulated sensorineural hearing loss. *Journal of Phonetics,* 28**,** 343-366.

LUTMAN, M. E. 23/03/2015 2015. *RE: Consent for Data Usage.* Type to BLYTH, M.

MACPHERSON, A. & AKEROYD, M. A. 2014. Variations in the slope of the psychometric functions for speech intelligibility: A systematic survey. *Trends in Hearing,* 18**,** 1-26.

MARKEL, N. N., PREBOR, L. D. & BRANDT, J. F. 1972. Biosocial factors in dyadic communication: Sex and speaking intensity. *Journal of Personality and Social Psychology,* 23**,** 11-13.

MATTHEWS, G., WARM, J. S., SHAW, T. H. & FINOMORE, V. S. 2014. Predicting battlefield vigilance: a multivariate approach to assessment of attentional resources. *Ergonomics,* 57**,** 856-875.

MCGILLVRAY, J. W. 1994. Stealth Technology in Surface Warships. *Naval War College Review,* 47**,** 28-39.

MCOLES. 2016. *Michigan Commission on Law Enforcement Standards. Employment Standards for Michigan Law Enforcement Officers* [Online]. http://www.michigan.gov/mcoles/. [Accessed 29/02/2016 2016].

MELE, M. L. & FEDERICI, S. 2012. Gaze and eye-tracking solutions for psychological research. *Cognitive processing,* 13**,** 261-265.

MENTEL, K., ZIRIAX, J., DACHOS, J., SALUNGA, A., TURNER, H., SHEFFIELD, B. & BRUNGART, D. The Relationship between Speech Intelligibility and Operational Performance in a Simulated Naval Command Information Center. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2013. SAGE Publications, 1204-1208.

MIDDELWEERD, M., FESTEN, J. & PLOMP, R. 1990. Difficulties with Speech Intelligibility in noise in spite of a normal pure-tone audiogram: original papers. *Audiology,* 29**,** 1-7.

MILLIGAN, G. S., REILLY, T. J., ZUMBO, B. D. & TIPTON, M. J. 2016. Validity and reliability of physical employment standards. *Applied Physiology, Nutrition, and Metabolism,* 41**,** S83-S91.

MOORE, B. C. 2004. Dead regions in the cochlea: conceptual foundations, diagnosis, and clinical applications. *Ear and hearing,* 25**,** 98-116.

MOORE, B. C. 2014. Development and current status of the “Cambridge” loudness models. *Trends in hearing,* 18**,** 1-29.

MOORE, B. C. & GLASBERG, B. R. 1993. Simulation of the effects of loudness recruitment and threshold elevation on the intelligibility of speech in quiet and in a background of speech. *The Journal of the Acoustical Society of America,* 94**,** 2050-2062.

MOORE, B. C., GLASBERG, B. R. & BAER, T. 1997. A model for the prediction of thresholds, loudness, and partial loudness. *Journal of the Audio Engineering Society,* 45**,** 224-240.

MOORE, B. C., GLASBERG, B. R., VARATHANATHAN, A. & SCHLITTENLACHER, J. 2016. A loudness model for time-varying sounds incorporating binaural inhibition. *Trends in hearing,* 20**,** 2331216516682698.

MOORE, B. C., GLASBERG, B. R. & VICKERS, D. A. 1995. Simulation of the effects of loudness recruitment on the intelligibility of speech in noise. *British Journal of Audiology,* 29**,** 131-143.

MOORE, B. C., VICKERS, D. A., BAER, T. & LAUNER, S. 1999. Factors affecting the loudness of modulated sounds. *The Journal of the Acoustical Society of America,* 105**,** 2757-2772.

MOORE, B. C. J. & GLASBERG, B. R. 2007. Modeling binaural loudness. *The Journal of the Acoustical Society of America,* 121**,** 1604-1612.

MURRAY, D. L., BOUTIN, S., O'DONOGHUE, M. & NAMS, V. O. 1995. Hunting behaviour of a sympatric felid and canid in relation to vegetative cover. *Animal Behaviour,* 50**,** 1203-1210.

MUSIEK, F. E., SHINN, J., CHERMAK, G. D. & BAMIOU, D.-E. 2017. Perspectives on the pure-tone audiogram. *Journal of the American Academy of Audiology,* 28**,** 655-671.

NAKASHIMA, A. & MCDAVID, K. 2018. Protecting the hearing-impaired worker: speech understanding with electronic hearing protection devices. *Journal of Military, Veteran and Family Health***,** 1-9.

NEJIME, Y. & MOORE, B. C. J. 1997. Simulation of the effect of threshold elevation and loudness recruitment combined with reduced frequency selectivity on the intelligibility of speech in noise. *The Journal of the Acoustical Society of America,* 102**,** 603-615.

NELSON, T., WALL, C., DRIVER, J. & SIMPSON, R. 2012. Op HERRICK primary care casualties: the forgotten many. *Journal of the Royal Army Medical Corps,* 158**,** 252-256.

O'LEARY, A. & WALLACH, H. 1980. Familiar size and linear perspective as distance cues in stereoscopic depth constancy. *Perception & Psychophysics,* 27**,** 131-135.

OKPALA, N. C. 2007. Knowledge and attitude of infantry soldiers to hearing conservation. *Military Medicine,* 172**,** 520-2.

ORGANIZATION, W. H. 1996. ISO 9613-2: Acoustics-Attenuation of sound propagation outdoors, Part 2: General method of calculation. *Geneva: International Organization for Standardization*.

PADDAN, G. S. & LOWER, M. C. 2016. Factors affecting sound exposure from firing an SA80 high-velocity rifle. *Applied Acoustics,* 109**,** 61-68.

PAPADOPOULOS, T., EDWARDS, D. S., ROWAN, D. & ALLEN, R. 2011. Identification of auditory cues utilized in human echolocation—Objective measurement results. *Biomedical Signal Processing and Control,* 6**,** 280-290.

PARSONS, T. D. 2015. Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in human neuroscience,* 9**,** 660.

PAVLOVIC, C. 2006. The speech intelligibility index standard and its relationship to the articulation index, and the speech transmission index. *The Journal of the Acoustical Society of America,* 119**,** 3326-3326.

PERROTT, D. R., SABERI, K., BROWN, K. & STRYBEL, T. Z. 1990. Auditory psychomotor coordination and visual search performance. *Perception & psychophysics,* 48**,** 214-226.

PETERS, L. J. & GARINTHER, G. R. 1990. The effects of speech intelligibility on crew performance in an M1A1 tank simulator. DTIC Document.

PEUGH, J. L. & ENDERS, C. K. 2004. Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of educational research,* 74**,** 525-556.

PHATAK, S. A., SHEFFIELD, B. M., BRUNGART, D. S. & GRANT, K. W. 2018. Development of a Test Battery for Evaluating Speech Perception in Complex Listening Environments: Effects of Sensorineural Hearing Loss. *Ear and hearing,* 39**,** 449-456.

PHILBECK, J. W. & MERSHON, D. H. 2002. Knowledge about typical source output influences perceived auditory distance. *The Journal of the Acoustical Society of America,* 111**,** 1980-1983.

PHILLIPS, D. P. 1999. Auditory gap detection, perceptual channels, and temporal resolution in speech perception. *JOURNAL-AMERICAN ACADEMY OF AUDIOLOGY,* 10**,** 343-354.

PICARD, M., BANVILLE, R., BARBAROSIE, T. & MANOLACHE, M. 1999. Speech audiometry in noise-exposed workers: the SRT-PTA relationship revisited. *Audiology,* 38**,** 30-43.

PIERCY, J. E., EMBLETON, T. F. W. & SUTHERLAND, L. C. 1977. Review of noise propagation in the atmosphere. *The Journal of the Acoustical Society of America,* 61**,** 1403-1418.

PLACK, C. J., BARKER, D. & PRENDERGAST, G. 2014. Perceptual consequences of “hidden” hearing loss. *Trends in hearing,* 18.

PLACK, C. J., LÉGER, A., PRENDERGAST, G., KLUK, K., GUEST, H. & MUNRO, K. J. 2016. Toward a diagnostic test for hidden hearing loss. *Trends in hearing,* 20**,** 1-9.

PLOMP, R. 1986. A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. *Journal of Speech, Language, and Hearing Research,* 29**,** 146-154.

POULSEN, T. 2010. Equivalent threshold sound pressure levels (ETSPL) for Interacoustics DD 45 supra-aural audiometric earphones. *International journal of audiology,* 49**,** 850-855.

POULSEN, T. & OAKLEY, S. 2009. Equivalent threshold sound pressure levels (ETSPL) for Sennheiser HDA 280 supra-aural audiometric earphones in the frequency range 125 Hz to 8000 Hz. *International Journal of Audiology,* 48**,** 271-276.

POWELL, R. & FORREST, M. 1988. *Noise in the military environment*, Brassey's Defence Publishers.

PREMACK, D. & WOODRUFF, G. 1978. Does the chimpanzee have a theory of mind? *Behavioral and brain sciences,* 1**,** 515-526.

PRICE, G. R., KALB, J. T. & GARINTHER, G. R. 1989. Toward a measure of auditory handicap in the Army. *The Annals of otology, rhinology & laryngology. Supplement,* 140**,** 42-52.

PRINS, N. & KINGDOM, F. A. 2016. *Psychophysics: a practical introduction*, Academic Press.

PRINS, N. & KINGDOM, F. A. 2018. Applying the model-comparison approach to test specific research hypotheses in psychophysical research using the Palamedes Toolbox. *Frontiers in psychology,* 9.

R. GARINTHER, G., T. KALB, J., C. HODGE, D. & RICHARD PRICE, G. 1985. *Proposed Aural Nondetectability Limits for Army Materiel*.

RENNER, R. S., VELICHKOVSKY, B. M. & HELMERT, J. R. 2013. The perception of egocentric distances in virtual environments-a review. *ACM Computing Surveys (CSUR),* 46**,** 23.

RENNIES, J., VERHEY, J. L., APPELL, J. E. & KOLLMEIER, B. 2013. Loudness of complex time-varying sounds ? A challenge for current loudness models. *Proceedings of Meetings on Acoustics,* 19**,** 1-9.

RICHARDSON, A. R. & WALLER, D. 2007. Interaction with an immersive virtual environment corrects users' distance estimates. *Human Factors,* 49**,** 507-517.

RIESER, J. J., ASHMEAD, D. H., TALOR, C. R. & YOUNGQUIST, G. A. 1990. Visual perception and the guidance of locomotion without vision to previously seen targets. *Perception,* 19**,** 675-689.

RONA, R., HOOPER, R., GREENBERG, N., JONES, M. & WESSELY, S. 2006. Medical downgrading, self-perception of health, and psychological symptoms in the British Armed Forces. *Occupational and environmental medicine,* 63**,** 250-254.

RONNBERG, J., RUDNER, M., LUNNER, T. & ZEKVELD, A. A. 2010. When cognition kicks in: Working memory and speech understanding in noise. *Noise & Health,* 12**,** 263-269.

ROVIG, G. W., BOHNKER, B. K. & PAGE, J. C. 2004. Hearing health risk in a population of aircraft carrier flight deck personnel. *Military medicine,* 169**,** 429-432.

ROWAN, D., PAPADOPOULOS, T., ARCHER, L., GOODHEW, A., COZENS, H., LOPEZ, R. G., EDWARDS, D., HOLMES, H. & ALLEN, R. 2017. The detection of 'virtual' objects using echoes by humans: Spectral cues. *Hearing Research,* 350**,** 205-216.

ROWAN, D. & PICKERING, R. 2011. Letter regarding two recent publications on interaural attenuation in the International Journal of Audiology. *International journal of audiology,* 50**,** 636-637.

ROY, S., MILLER, C. T., GOTTSCH, D. & WANG, X. 2011. Vocal control by the common marmoset in the presence of interfering noise. *Journal of Experimental Biology,* 214**,** 3619-3629.

RUBIN, D. B. 1976. Inference and missing data. *Biometrika,* 63**,** 581-592.

RUDZYN, B. & FISHER, M. 2012. Performance of personal active noise reduction devices. *Applied Acoustics,* 73**,** 1159-1167.

RYDELL, J., SKALS, N., SURLYKKE, A. & SVENSSON, M. 1997. Hearing and bat defence in geometrid winter moths. *Proceedings of the Royal Society of London. Series B: Biological Sciences,* 264**,** 83-88.

SCHAFER, J. L. & GRAHAM, J. W. 2002. Missing data: our view of the state of the art. *Psychological methods,* 7**,** 147.

SCHARINE, A., LETOWSKI, T. & SAMPSON, J. B. 2009. Auditory situation awareness in urban operations. *Journal of Military and Strategic Studies,* 11.

SCHOOF, T. & ROSEN, S. 2014. The role of auditory and cognitive factors in understanding speech in noise by normal-hearing older listeners. *Front Aging Neurosci,* 6**,** 307.

SEEBACHER, F., WARD, A. & REID, A. 2010. Learning to hunt: the role of experience in predator success. *Behaviour,* 147**,** 223-233.

SEMERARO, H. 2016. *Developing a measure of auditory fitness for duty for military personnel.* University of Southampton.

SEMERARO, H. D., BEVIS, Z. L., ROWAN, D., VAN BESOUW, R. M. & ALLSOPP, A. J. 2015. Fit for the frontline? Identification of mission-critical auditory tasks (MCATs) carried out by infantry and combat-support personnel. *Noise and Health,* 17**,** 98-107.

SEYMOUR, N. E., GALLAGHER, A. G., ROMAN, S. A., O'BRIEN, M. K., BANSAL, V. K., ANDERSEN, D. K. & SATAVA, R. M. 2002. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Annals of surgery,* 236**,** 458.

SHAPIRO, S. S. & WILK, M. B. 1965. An analysis of variance test for normality (complete samples). *Biometrika,* 52**,** 591-611.

SHEFFIELD, B., BRUNGART, D. & BLANK, A. The Effects of Hearing Impairment on Fire Team Performance in Dismounted Combat. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2016a. SAGE Publications Sage CA: Los Angeles, CA, 1509-1513.

SHEFFIELD, B., BRUNGART, D., TUFTS, J. & NESS, C. J. The relationship between hearing acuity and operational performance in dismounted combat. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2015. SAGE Publications Sage CA: Los Angeles, CA, 1346-1350.

SHEFFIELD, B., BRUNGART, D., TUFTS, J. & NESS, J. 2016b. The effects of elevated hearing thresholds on performance in a paintball simulation of individual dismounted combat. *International Journal of Audiology***,** 1-7.

SHEFFIELD, B., BRUNGART, D., TUFTS, J. & NESS, J. 2017. The effects of elevated hearing thresholds on performance in a paintball simulation of individual dismounted combat. *International journal of audiology,* 56**,** 34-40.

SHINN-CUNNINGHAM, B. G. 2008. Object-based auditory and visual attention. *Trends in Cognitive Science,* 12**,** 182-6.

SHINN-CUNNINGHAM, B. G., IHLEFELD, A., SATYAVARTA & LARSON, E. 2005. Bottom-up and top-down influences on spatial unmasking. *Acta Acustica United with Acustica,* 91**,** 967-979.

SMOORENBURG, G. F. 1992. Speech reception in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiogram. *Journal of the Acoustical Society of America,* 91**,** 421-37.

SOLI, S. D., AMANO-KUSUMOTO, A., CLAVIER, O., WILBUR, J., CASTO, K., FREED, D., LAROCHE, C., VAILLANCOURT, V., GIGUÈRE, C. & DRESCHLER, W. A. 2018a. Evidence-based occupational hearing screening II: validation of a screening methodology using measures of functional hearing ability. *International journal of audiology,* 57**,** 323-334.

SOLI, S. D., GIGUÈRE, C., LAROCHE, C., VAILLANCOURT, V., DRESCHLER, W. A., RHEBERGEN, K. S., HARKINS, K., RUCKSTUHL, M., RAMULU, P. & MEYERS, L. S. 2018b. Evidence-based occupational hearing Screening I: Modeling the effects of real-world noise environments on the likelihood of effective speech communication. *Ear and hearing,* 39**,** 436-448.

STONE, M. A. & MOORE, B. C. 1999. Tolerable hearing aid delays. I. Estimation of limits imposed by the auditory path alone using simulated hearing losses. *Ear and Hearing,* 20**,** 182-192.

STRASBURGER, H. 2001. Converting between measures of slope of the psychometric function. *Perception & psychophysics,* 63**,** 1348-1355.

SUMMERFIELD, Q., PALMER, A., FOSTER, J., MARSHALL, D. & TWOMEY, T. 1994. Clinical evaluation and test-retest reliability of the IHR-McCormick Automated Toy Discrimination Test. *British journal of audiology,* 28**,** 165-179.

SURPRENANT, A. M. & WATSON, C. S. 2001. Individual differences in the processing of speech and nonspeech sounds by normal-hearing listeners. *The Journal of the Acoustical Society of America,* 110**,** 2085-2095.

TANENHOLTZ, S. D. 1968. Research on acoustical problems of the military: A review and future aspect. ARMY NATICK LABS MA PIONEERING RESEARCH LAB.

TARR, M. J. & WARREN, W. H. 2002. Virtual reality in behavioral neuroscience and beyond. *nature neuroscience,* 5**,** 1089.

TECHNOLOGIES, U. 2016. Unity 5.5. *In:* TECHNOLOGIES, U. (ed.).

TECHNOLOGIES, U. 2017. Unity. 5.5.4 ed.

TEPE, V., SMALT, C., NELSON, J., QUATIERI, T. & PITTS, K. 2017. Hidden hearing injury: The emerging science and military relevance of cochlear synaptopathy. *Military medicine,* 182**,** e1785-e1795.

TER KEURS, M., FESTEN, J. M. & PLOMP, R. 1989. Effects of spectral smearing on speech perception. *The Journal of the Acoustical Society of America,* 86**,** S80-S80.

TREMBLAY, K. L., PINTO, A., FISCHER, M. E., KLEIN, B. E., KLEIN, R., LEVY, S., TWEED, T. S. & CRUICKSHANKS, K. J. 2015. Self-Reported Hearing Difficulties Among Adults With Normal Audiograms: The Beaver Dam Offspring Study. *Ear Hear,* 36**,** e290-e299.

TROST, R. P. & SHAW, G. B. 2007. Statistical analysis of hearing loss among navy personnel. *Military medicine,* 172**,** 426-430.

TUFTS, J. B., AUDETTE, W., WILBUR, J. & CLAVIER, O. 2018. Development and implementation of a military auditory fitness-for-duty test battery. *The Journal of the Acoustical Society of America,* 143**,** 1780-1780.

TUFTS, J. B., VASIL, K. A. & BRIGGS, S. 2009. Auditory fitness for duty: a review. *Journal of the American Academy of Audiology,* 20**,** 539-557.

TUFTS, J. B., WEATHERSBY, P. K. & RODRIGUEZ, F. A. 2010. Modeling the Unites States government's economic cost of noise-induced hearing loss for a military population. *Scandinavian journal of work, environment & health***,** 242-249.

TVERSKY, A. & KAHNEMAN, D. 1974. Judgment under uncertainty: Heuristics and biases. *science,* 185**,** 1124-1131.

VILLCHUR, E. 1974. Simulation of the effect of recruitment on loudness relationships in speech. *The Journal of the Acoustical Society of America,* 56**,** 1601-1611.

WEN, G., LU, Y., ZHANG, Z., MA, C., YIN, H. & CUI, Z. 2009. Line spectra reduction and vibration isolation via modified projective synchronization for acoustic stealth of submarines. *Journal of Sound and Vibration,* 324**,** 954-961.

WENZEL, E. M., ARRUDA, M., KISTLER, D. J. & WIGHTMAN, F. L. 1993. Localization using nonindividualized head-related transfer functions. *The Journal of the Acoustical Society of America,* 94**,** 111-123.

WHITTLE, L. & DELANY, M. 1966. Equivalent Threshold Sound-Pressure Levels for the TDH39/MX41-AR Earphone. *The Journal of the Acoustical Society of America,* 39**,** 1187-1188.

WICHMANN, F. A. & HILL, N. J. 2001. The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & psychophysics,* 63**,** 1293-1313.

WIENER, F. M. & KEAST, D. N. 1959. Experimental study of the propagation of sound over ground. *The Journal of the Acoustical Society of America,* 31**,** 724-733.

WIGNALL, A. E., JACKSON, R. R., WILCOX, R. S. & TAYLOR, P. W. 2011. Exploitation of environmental noise by an araneophagic assassin bug. *Animal Behaviour,* 82**,** 1037-1042.

WILEY, T. L., CRUICKSHANKS, K. J., CRUICKHANKS, K. J., NONDAHL, D. M. & TWEED, T. S. 2000. Self-reported hearing handicap and audiometric measures in older adults. *Journal of the American Academy of Audiology,* 11.

WILSON, R. H. & CARTER, A. S. 2001. Relation between slopes of word recognition psychometric functions and homogeneity of the stimulus materials. *Journal of the American Academy of Audiology,* 12**,** 7-14.

WITMER, B. G. & KLINE, P. B. 1998. Judging perceived and traversed distance in virtual environments. *Presence,* 7**,** 144-167.

WRANGHAM, R., WILSON, M. & HAUSER, M. 2007. Chimpanzees (Pan troglodytes) modify grouping and vocal behaviour in response to location-specific risk. *Behaviour,* 144**,** 1621-1653.

YACK, J. E. & FULLARD, J. H. 2000. Ultrasonic hearing in nocturnal butterflies. *Nature,* 403**,** 265.

YANKASKAS, K. 2013. Prelude: Noise-induced tinnitus and hearing loss in the military. *Hearing Research,* 295**,** 3-8.

ZAHORIK, P. 2002. Assessing auditory distance perception using virtual acoustics. *The Journal of the Acoustical Society of America,* 111**,** 1832-1846.

ZAHORIK, P., BRUNGART, D. S. & BRONKHORST, A. W. 2005. Auditory distance perception in humans: A summary of past and present research. *Acta Acustica united with Acustica,* 91**,** 409-420.

ZAHORIK, P. & KELLY, J. W. 2007. Accurate vocal compensation for sound intensity loss with increasing distance in natural environments. *The Journal of the Acoustical Society of America,* 122**,** EL143-EL150.

ZAHORIK, P. & WIGHTMAN, F. L. 2001. Loudness constancy with varying sound source distance. *Nat Neurosci,* 4**,** 78-83.

ZEKVELD, A. A., RUDNER, M., JOHNSRUDE, I. S., FESTEN, J. M., VAN BEEK, J. H. & RONNBERG, J. 2011. The influence of semantically related and unrelated text cues on the intelligibility of sentences in noise. *Ear Hear,* 32**,** e16-25.

ZOLLINGER, S. A. & BRUMM, H. 2011. The Lombard effect. *Current Biology,* 21**,** R614-R615.

ZORILĂ, T.-C., STYLIANOU, Y., FLANAGAN, S. & MOORE, B. C. J. 2016. Effectiveness of a loudness model for time-varying sounds in equating the loudness of sentences subjected to different forms of signal processing. *The Journal of the Acoustical Society of America,* 140**,** 402-408.